Lecture Notes for Theory of Value: EC 121a

Farzad Pourbabaee
Division of the Humanities and Social Sciences – Caltech

November 29, 2021

Contents

Acknowledgements

These notes are majorly based on the lectures of late Kim C. Border. Students and readers of these notes are encouraged to regularly check his website for further insights about mathematical economics.

Lecture 1

Introduction to Comparative Statics

The models we study in this class are largely **qualitative**. Here is a simple example of what I mean by a qualitative model.

1.1 Supply and Demand in a Single Market

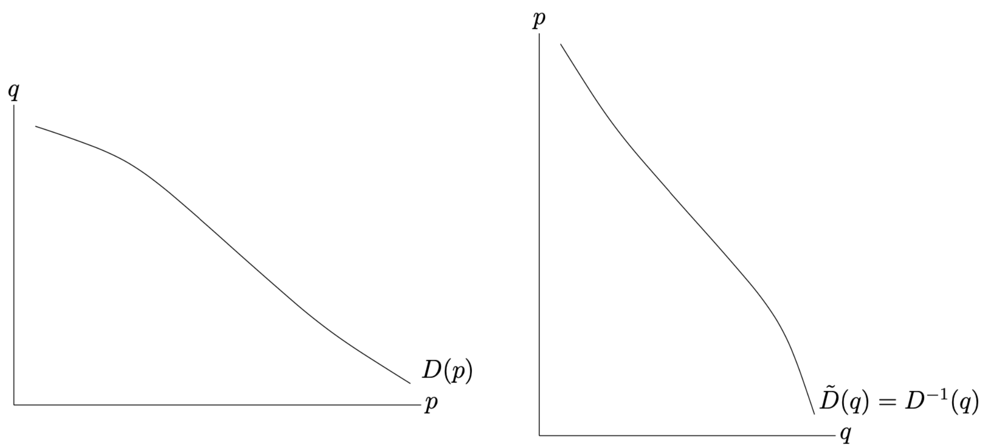
You may recall from EC 11 that a **demand curve** expresses a relation between quantities and prices for buyers. This being economics, there are at least two interpretations of the demand curve. The **Walrasian demand curve** gives the quantity that buyers are willing to buy as a function of the price they pay. As a function it maps prices to quantities. The **Marshallian demand curve** gives for each quantity the price at which buyers are willing to buy that quantity (see figure 1.1).

1.2 Comparative Statics of Supply and Demand

The model has these pieces: demand, supply and the market clearing price.

The (Walrasian) **demand curve** gives the quantity that buyers are willing to buy as a function of the price they pay, $D(p)$ is the quantity demanded at price p. As a behavioral assumption, we normally expect the demand curve to be downward sloping. To simplify things, let us assume the demand curve is smooth and that $D'(p) < 0$.

The (Walrasian) **supply curve** give the quantity that sellers are willing to sell as a function of the price they receive, namely $S(p)$ is the quantity supplied at price p. We normally expect the supply curve to be upward sloping and further is smooth, that is $S'(p) > 0$ (see figure 1.2).



(a) A (Walrasian) demand curve

(b) The same curve as a Marshallian demand

Figure 1.1: Demand curves

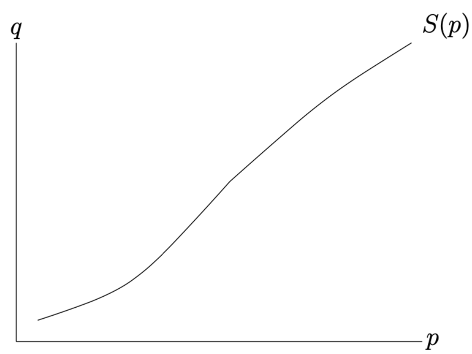


Figure 1.2: A (Walrasian) supply curve

The **market equilibrium price** p^* and **equilibrium quantity** q^* are determined by "market forces" so that the quantity demanded is equal to the quantity supplied, or the market "**clears**." That is,

$$D(p^*) - S(p^*) = 0, \tag{1.1}$$

$$q^* = D(p^*) = S(p^*). \tag{1.2}$$

What does **comparative statics** mean? The testable content does not come from characterizing the price-quantity pairs that can be market equilibria, but rather from the way the

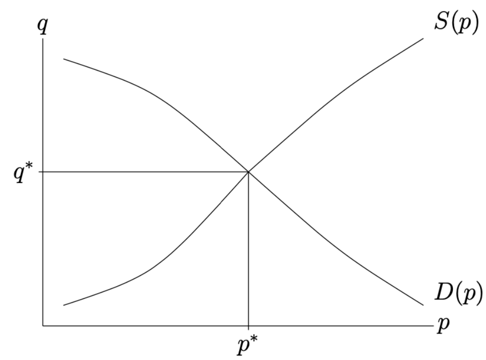


Figure 1.3: Market clearing

equilibrium changes in response to *interventions* or exogenous changes to the environment. The identification of how the static equilibrium changes in response to changes in outside factors is called **comparative statics**.

Example 1.1 (Testable implications)**.** Suppose a tax of t per unit sold is imposed. What happens to the prices?

To answer this we need to be careful about which price we are speaking of. Let p_b denote the price the buyer pays and p_s the price the seller receives. Then $p_b = p_s + t$. The market prices $p_b(t)$ and $p_s(t)$, which depend on the size of the tax, are determined by "market forces" so that the quantity demanded is equal to the quantity supplied.

$$\begin{gathered} D(p_b(t)) - S(p_s(t)) = 0, \\ q(t) = D(p_b(t)) = S(p_s(t)), \\ p_b(t) = p_s(t) + t. \end{gathered} \tag{1.3}$$

We are interested in how the equilibrium prices $p_b(t)$ and $p_s(t)$ vary with the tax t. In other words, we want to know what we can say about p_s', p_b', and q' (which are the derivatives with respect ot t). For starters, if there are market clearing prices, we know that for all t, $D(p_s(t) + t) - S(p_s(t)) = 0$. Differentiate both sides with respect to t leads to:

$$\begin{gathered} D'(p_s(t) + t)(p_s'(t) + 1) - S'(p_s(t))p_s'(t) = 0, \\ \Rightarrow p_s'(t) = -\frac{D'(p_s(t) + t)}{D'(p_s(t) + t) - S'(p_s(t))}. \end{gathered} \tag{1.4}$$

Since $D' < 0$ and $S' > 0$, then it follows that

$$-1 < p_s'(t) < 0. \tag{1.5}$$

A similar analysis implies that $0 < p_b'(t) < 1$. Jointly these two relations imply that as the tax rate increases, the equilibrium price that buyer pays increases and the equilibrium price the seller receives decreases.

Exercise 1.2. In the previous example, perform the comparative statics of equilibrium amount $q(t)$ with respect to the tax t.

How do we know that market clearing prices will exist? That is, how can we be sure that the supply and demand curves cross? Here is a sufficient set of conditions that will guarantee it (assuming smoothness, which implies no jumps). If the price is low enough, then demand exceeds supply.

$$\lim_{p\to 0} D(p) > \lim_{p\to 0} S(p).$$

If the price is high enough, then supply exceeds demand:

$$\lim_{p\to\infty} S(p) > \lim_{p\to\infty} D(p).$$

Exercise 1.3. In the above comparative statics example we took the derivatives of equilibrium values $p_b(t)$ and $p_s(t)$ with respect to t without proving these functions are differentiable. Prove that this is indeed the case.

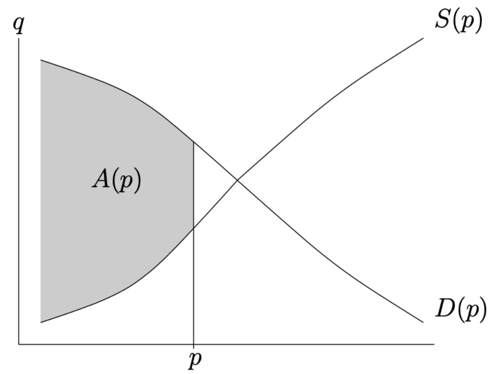
1.3 Market equilibria as maximizers

For a smooth function f, if its derivative is positive to the left of x^* and negative to the right, then it achieves a maximum at x^*. In our simple market, demand is greater than supply to the left of equilibrium and less than supply to the right. Thus excess demand (demand – supply) acts like the derivative of a function that has a maximum at the market equilibrium.

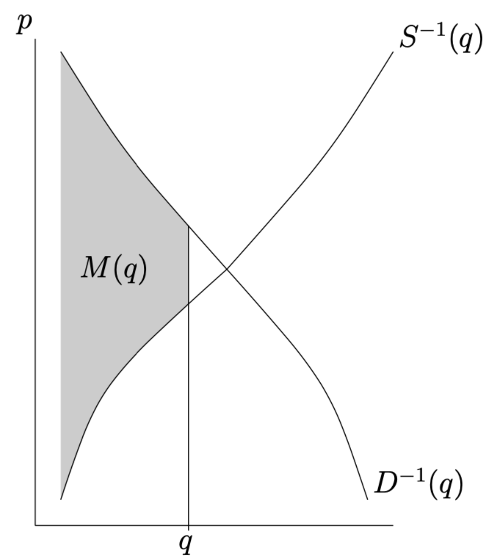
Define $A(p)$ by

$$A(p) = \int_0^p (D(x) - S(x))\,\mathrm{d}x,$$

which is the (signed) area under the Walrasian demand curve and above the Walrasian supply curve up to p (see figure 1.4a). There is a mathematical chance that this area might



(a) Market equilibrium maximizes $A(p)$



(b) Market equilibrium maximizes Marshallian surplus $M(q)$

Figure 1.4: Equilibrium as maximizers

be infinite, but that could not be the case for a demand curve, as it would imply that the revenue obtainable by lowering the price and increasing the quantity would be unbounded, which cannot happen in a real economy. Then, by the First Fundamental Theorem of Calculus

$$A'(p) = D(p) - S(p), \text{and} A''(p) = D'(p) - S'(p) < 0,$$

so A is strictly concave. Its maximum occurs at the market clearing price p^*. Thus, at least in this simple case, *finding the market clearing price is equivalent to maximizing an appropriate function.* This is a theme to which we shall return later. But now let us recast the above argument in a Marshallian framework.

The Marshallian demand curve is the inverse of the Walrasian demand curve and ditto for the supply curves. Marshallian **consumers' surplus** is the area under the Marshallian demand up to some quantity q. You may recall from Ec 11 (we shall derive it later) that the inverse supply curve of a price-taking seller is the marginal cost. Thus the area under the inverse supply curve up to q is the total variable cost of producing q. Thus the (signed) area $M(q)$ between the inverse demand curve and the inverse supply curve is equal to consumers' surplus minus (variable) cost. Once again, the market equilibrium quantity q^* maximizes the **Marshallian surplus** $M(q)$. See figure 1.4b.

Appendix: The Implicit Function Theorem

An equation of the form

$$f(x,p) = y \tag{1.6}$$

implicitly defines x **as a function of** $p$on a domain P if there is a function ξ on P for which

$$f(\xi(p),p) = y$$

for all $p \in P$. It is traditional to assume that $y = 0$, but not essential. (We can always convert y to zero by defining $\hat{f}(x,p) = f(x,p) - y$. Then $f(x,p) = y$ if and only if $\hat{f}(x,p) = 0$.)

Theorem 1.4 (Classical Implicit Function Theorem)**.** *Let $X \times P$ be an open subset of $\mathbb{R}^n \times \mathbb{R}^m$, and let $f : X \times P \to \mathbb{R}^n$ be C^k, for $k \geq 1$. Assume that $D_x f(\bar{x},\bar{p})$ is invertible.*

Then there are neighborhoods $U \subset X$ and $W \subset P$ of $\bar{x}$ and $\bar{p}$ on which (1.6) *uniquely defines x as a function of p. That is, there is a function $\xi : W \to U$ such that:*

1. *$f(\xi(p);p) = f(\bar{x},\bar{p})$ for all $p \in W$.*
2. *For each $p \in W$, $\xi(p)$ is the unique solution to* (1.6) *lying in U. In particular, then $\xi(\bar{p}) = \bar{x}$.*
3. *ξ is C^k on W, and*

$$\begin{bmatrix} \frac{\partial \xi_1}{\partial p_1} & \cdots & \frac{\partial \xi_1}{\partial p_m} \\ \vdots & & \vdots \\ \frac{\partial \xi_n}{\partial p_1} & \cdots & \frac{\partial \xi_n}{\partial p_m} \end{bmatrix} = - \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial f_1}{\partial p_1} & \cdots & \frac{\partial f_1}{\partial p_m} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial p_1} & \cdots & \frac{\partial f_n}{\partial p_m} \end{bmatrix}. \tag{1.7}$$

We end this chapter with few examples in which the Implicit Function Theorem fails to hold.

Example 1.5 (Differential not invertible)**.** Define $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by

$$f(x,p) = (x - p^2)(x - 2p^2).$$

Consider the function implicitly defined by $f(x,p) = 0$. The function f is zero along the parabolas $x = p^2$ and $x = 2p^2$, and in particular $f(0,0) = 0$. The hypothesis of the Implicit Function Theorem is not satisfied since $\frac{\partial f(0,0)}{\partial x} = 0$. The conclusion also fails. The problem here is not that a smooth implicit function through $(x,p) = (0,0)$ fails to exist. The problem is that it is not unique. There are four distinct continuously differentiable implicitly defined functions.

Example 1.6 (Lack of continuous differentiability)**.** Consider the function $h(x) = x + 2x^2 \sin \frac{1}{x^2}$, see figure 1.5. This function is differentiable everywhere, but not continuously differentiable at zero. Furthermore, $h(0) = 0$, $h'(0) = 1$, but h is not monotone on any neighborhood of zero. Now consider the function $f(x, p) = h(x) - p$. It satisfies $f(0, 0) = 0$ and $\frac{\partial f(0,0)}{\partial x} \neq 0$, but there is no unique implicitly defined function on any neighborhood, nor is there any continuous implicitly defined function.

To see this, note that $f(x, p) = 0$ if and only if $h(x) = p$. So a unique implicitly defined function exists only if h is invertible on some neighborhood of zero. But this is not so, for given any $\epsilon > 0$, there is some $0 < p < \frac{\epsilon}{2}$ for which there are $0 < x < x' < \epsilon$ satisfying $h(x) = h(x') = p$. It is also easy to see that no continuous function satisfies $h\big(\xi(p)\big) = p$ either.

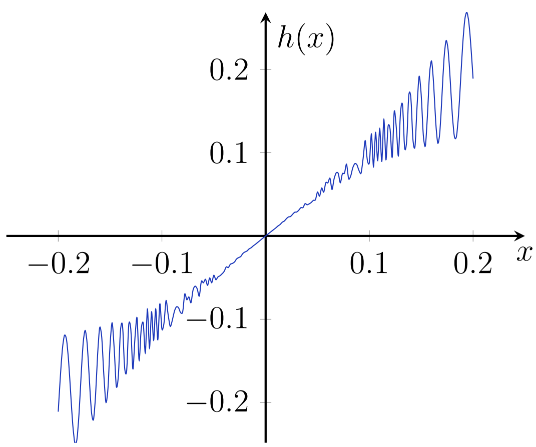


Figure 1.5: $h(x) = x + 2x^2 \sin \frac{1}{x^2}$

Lecture 2

Profit Maximization

2.1 Maximization and comparative statics

Just as above, our "equilibrium" conditions are often the results of some maximizing behavior. Consider this simple model of a firm. When the firm produces the level $y \geq 0$ of output, it receives revenue $R(y)$ and incurs cost $C(y)$. The profit is then $R(y) - C(y)$. In addition, it pays an *ad rem* tax ty. It seeks to maximize its after-tax profit:

$$\max \; R(y) - C(y) - ty.$$

Let $y^*(t)$ solve this problem. What do we know?

$$R'(y^*) - C'(y^*) - t = 0$$

This does not tell us much about data that we might observe, but let's see how y^* changes with t:

$$R'(y^*(t)) - C'(y^*(t)) - t = 0 \quad \text{for all } t.$$

Therefore, by differentiating both sides with respect to t we get

$$[R''(y^*(t)) - C''(y^*(t))]\, y^{*\prime}(t) - 1 = 0,$$

or

$$y^{*\prime}(t) = \frac{1}{R''(y^*(t)) - C''(y^*(t))}.$$

How can we sign this? The answer is, via the second order conditions. Namely, $R''(y^*) - C''(y^*) \leq 0$, which implies $y^{*\prime}(t) < 0$. The Implicit Function Theorem guarantees that if $R''(y^*) - C''(y^*) \neq 0$, then $y^*(t)$ is unique and differentiable.

Revenue maximization

What if the firm maximizes after-tax revenue $R(y) - ty$ instead of profit. The first order condition is $R'(y) - t = 0$ and the second order condition is $R''(y) \leq 0$ (note that I have used the economists' sloppy notation of omitting the *. I should actually use something different, since it is a different function). Differentiating the first order condition with respect to t yields

$$R''(y)y' - 1 = 0,$$

or

$$y' = \frac{1}{R''(y)} < 0,$$

where the inequality follows from the strict second order condition. Thus a change in an *ad rem* tax gives us no leverage on deciding whether a firm after-tax maximizes revenue or after-tax profit.

Wages

Suppose that the firm's costs C are a function both of its level of output and a wage parameter, and assume that the partial derivative $D_yC > 0$ (which must be the case if the firm is minimizing costs).

For profit maximization,

$$\text{maximize } R(y) - C(y; w),$$

the first order condition is $R'(y) - D_yC(y; w) = 0$, for an interior solution (marginal cost = marginal revenue), and the second order condition is

$$R''(y) - D_y^2C(y; w) \leq 0.$$

Letting $y^*(w)$ be the maximizer we see that

$$h(w) = R'(y^*(w)) - D_yC(y^*(w); w) = 0,$$

for all w. Thus h is constant so $h' = 0$. By the chain rule,

$$h'(w) = R''(y^*(w))y^{*\prime}(w) - D_y^2 C(y^*(w); w)y^{*\prime}(w) - D_{yw}C(y^*(w); w) = 0.$$

Solving for $y^{*\prime}$ gives

$$y^{*\prime}(w) = \frac{D_{yw}C(y^*(w); w)}{R''(y^*(w)) - D_y^2 C(y^*(w); w)}.$$

The denominator must be negative, so the sign of this is the opposite of the sign of the mixed partial $D_{yw}C$.

For revenue maximization, i.e

$$\max R(y),$$

the first order condition is $R'(y) = 0$ for an interior solution, and the second order condition is $R''(y) \leq 0$. Letting $\hat{y}(w)$ be the maximizer we see that it is independent of w! Thus $\hat{y}'(w) = 0$!

An application to sports economics

What is the effect of player salaries on ticket prices?

For a profit-maximizing sports franchise (and one visit to Dodger Stadium ought to convince you that profits are being fiercely pursued), the revenue comes from ticket sales, parking, and concessions, but the costs are almost entirely determined by players' (and coaches' and groundskeepers') wages and utility bills for the lights, all of which do not depend on how many tickets are sold. The number of tickets sold will depend on the price charged, so the revenue is not going to be a linear function of the number of tickets sold.

A reasonable approximation to profit is

$$\text{profit} = R(y) - C(w),$$

where y is the number of tickets sold. (Parking and concessions tend to be proportional to the number of tickets.) There is also TV revenue, which does not depend on y, but can be treated as an additive constant. While some costs (free bobble heads, programs, etc.) are proportional to tickets they are small, and could also be netted out of the ticket revenues. We can see that w has no effect on y, so it cannot affect the ticket price.

What about second-order effects—higher wages attract better players, and so increase demand for tickets, enabling the franchise to sell the same number of tickets at a higher price. This works only if higher wages are limited to one team that is able to attract all the good players—if all teams' wages go up, there is no reason to expect any one team to get better.

What about third-order effects—if wages are too high the team will fold and then there will be no tickets available at any price. This might be more convincing if team prices were lower, but according to Forbes, as of September 2020, NFL franchises were worth on average over \$3 billion, ranging from from \$2 billion (Buffalo Bills) to \$6.5 billion (Dallas Cowboys). (See Forbes' list).

If, as the owners usually claim come time to negotiate with players, teams are such money losers, then why are team prices so high? For one thing, teams are a good tax shelter. A new owner can assign 80% of the value to player contracts and depreciate them over four or five years, then resell the team for largely capital gains. They are also frequently real cash cows. And then there are some special accounting practices that allow profits to be counted as expenses.

2.2 After tax profit revisited

$$\max R(y) - C(y) - ty.$$

Let y^* solve this problem. Last time we used the second order conditions to conclude

$$\frac{\mathrm{d}}{\mathrm{d}t} y^*(t) < 0,$$

provided the derivative exists.

But we got stuck when it came to dealing with wages. In that case

$$\operatorname{sgn}\left(\frac{\mathrm{d}}{\mathrm{d}w}\hat{y}(w)\right) = -\operatorname{sgn}\left(D_{yw}C(y,w)\right).$$

For this we can use another approach.

2.3 A lemma

Proposition 2.1. *Let X and P be open intervals in $\mathbb{R}$, and let $f : X \times P \to \mathbb{R}$ be twice continuously differentiable. Assume that for all $x \in X$ and all $p \in P$,*

$$\frac{\partial^2 f(x,p)}{\partial p \partial x} \geq 0. \tag{2.1}$$

Let x^0 maximize $f(\cdot, p^0)$ over X and x^1 maximize $f(\cdot, p^1)$ over X. Then

$$(p^1 - p^0)(x^1 - x^0) \geq 0. \tag{2.2}$$

In other words, the sign of the change in the maximizing x is the same as the sign of the change in p.

If $\leq$ replaces $\geq$ in (2.1), then the sign of the change in x is the opposite of the sign of the change in p.

For minimization rather than maximization the sign of the effect is reversed.

Proof. By definition of maximization, we have

$$f(x^0, p^0) \geq f(x^1, p^0) \quad \text{and} \quad f(x^1, p^1) \geq f(x^0, p^1).$$

"Cross-subtracting" implies

$$f(x^1, p^1) - f(x^1, p^0) \geq f(x^0, p^1) - f(x^0, p^0). \tag{$*$}$$

But

$$f(x^1, p^1) - f(x^1, p^0) = \int_{p_0}^{p^1} \frac{\partial f}{\partial p}(x^1, \pi)\, \mathrm{d}\pi$$

and

$$f(x^0, p^1) - f(x^0, p^0) = \int_{p^0}^{p^1} \frac{\partial f}{\partial p}(x^0, \pi)\, \mathrm{d}\pi.$$

So $(*)$ becomes

$$\int_{p^0}^{p^1} \frac{\partial f}{\partial p}(x^1, \pi)\, \mathrm{d}\pi \geq \int_{p^0}^{p^1} \frac{\partial f}{\partial p}(x^0, \pi)\, \mathrm{d}\pi,$$

or

$$\int_{p^0}^{p^1} \left(\frac{\partial f}{\partial p}(x^1, \pi) - \frac{\partial f}{\partial p}(x^0, \pi) \right) \mathrm{d}\pi \geq 0.$$

Now we use the same trick of writing a difference as an integral of the derivative to get

$$\int_{p^0}^{p^1} \left(\frac{\partial f}{\partial p}(x^1, \pi) - \frac{\partial f}{\partial p}(x^0, \pi) \right) \mathrm{d}\pi = \int_{p^0}^{p^1} \left(\int_{x^0}^{x^1} \frac{\partial^2 f}{\partial p \partial x}(\xi, \pi)\, \mathrm{d}\xi \right) \mathrm{d}\pi \geq 0.$$

By assumption $\frac{\partial^2 f}{\partial p \partial x} \geq 0$, so by the convention that $\int_a^b = -\int_b^a$, we conclude that if $p_1 > p_0$, then $x_1 \geq x_0$, and the conclusion follows. □

Application 1

So consider

$$f(y,t) = R(y) - C(y) - ty.$$

Then

$$\frac{\partial f(y,t)}{\partial t} = -y$$

so

$$\frac{\partial^2 f(y,t)}{\partial y \partial t} = -1 < 0,$$

so

$$\frac{d}{dt} y^*(t) < 0.$$

Exercise 2.2. Apply the above analysis to revenue maximization.

Exercise 2.3. Consider $y = f(x)$, where x is an input that gets paid wage w. The profit maximization problem is

$$\max pf(x) - wx.$$

Perform the comparative statics of x^* (optimal solution to the above problem) with respect to the output price p.

2.4 Supermodularity

If inequality $*$ holds whenever $x_1 > x_0$ and $p_1 > p_0$, we say that f exhibits **increasing differences**, a property related to what we now call **supermodularity**. To define this, we first need to define a **lattice**.

Definition 2.4. A **partial order** $\succeq$ on a set X is a binary relation that is transitive, reflexive, and antisymmetric. A **lattice** is a partially ordered set $(X, \succeq)$ with the property that every pair $x, y \in X$, has a least upper bound $x \vee y$ (also called the **join**) and a greatest lower bound $x \wedge y$ (also called the **meet**).

For now, the most important example of a lattice is $\mathbb{R}^n$ with the coordinate-wise ordering $\geqq$, where $x \geqq y$ if $x_i \geq y_i$ for each $i = 1, \dots, n$.

Definition 2.5. A real-valued function f on a lattice is **supermodular** if

$$f(x \wedge y) + f(x \vee y) \geq f(x) + f(y).$$

Proposition 2.1 can be restated as follows.

Proposition 2.6. *If f is a twice differentiable function on $(\mathbb{R}^n, \geqq)$, then f is supermodular if and only if for $i \neq j$*

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \geq 0.$$

That is, an increase in i increases the marginal increase of j. That is i and j are **complements**.

Lecture 3

Maximization and Comparative Statics

3.1 Maximization with many variables

Theorem 3.1 (Necessary First Order Conditions)**.** *If U is an open subset of a normed space, and $x^* \in U$ is a local extremum of f, and f has directional derivatives at x^*, then for any nonzero v, the directional derivative satisfies $D_v f(x^*) = 0$. In particular, if f is differentiable at x^*, then $Df(x^*) = 0$.*

Proof using one-dimensional case. Since x^* is an interior point of U, there is an $\varepsilon > 0$ such that $x^* + \lambda v \in U$ for any $\lambda \in (-\varepsilon, \varepsilon)$ and any $v \in \mathbb{R}^n$ with $|v| = 1$. Set $g_v(\lambda) = f(x^* + \lambda v)$. Then g_v has an extremum at $\lambda = 0$. Therefore $g_v'(0) = 0$. By the chain rule, $g_v'(\lambda) = Df(x^* + \lambda v)(v)$. Thus we see that $Df(x^*)(v) = 0$ for every v, so $Df(x^*) = 0$. □

Theorem 3.2 (Necessary second order conditions)**.** *Let f be a continuously differentiable real-valued function on an open subset U of $\mathbb{R}^n$ and assume that f is twice differentiable at x^*, and define the quadratic form $Q(v) = D^2 f(x^*)(v, v)$. If x^* is a local maximizer, then Q is negative semidefinite. If x^* is a local minimizer, then Q is positive semidefinite.*

Proof using the chain rule. As in the proof of Theorem 3.1, define $g(\lambda) = f(x^* + \lambda v)$. it achieves a maximum at $\lambda = 0$, so the second order condition is $g''(0) \leq 0$. So by the chain rule, using $Df(x^*) = 0$,

$$g''(0) = D^2 f(x^*)(v, v) \leq 0.$$

That is, Q is negative semidefinite. □

Theorem 3.3 (Sufficient second order conditions)**.** *Let f be a continuously differentiable real-valued function on an open subset U of $\mathbb{R}^n$. Let x^* belong to U and assume that $Df(x^*) = 0$ and that f is twice differentiable at x^*.*

If the Hessian matrix $f''(x^*)$ is positive definite, then x^* is a strict local minimizer of f.

If the Hessian matrix $f''(x^*)$ is negative definite, then x^* is a strict local maximizer of f.

If the Hessian is nonsingular but indefinite, then x^* is neither a local maximum, nor a local minimum.

3.2 Comparative statics of first order conditions

Start with a function $f : X \times P \to \mathbb{R}$ where $X \subset \mathbb{R}^n$ and $P \subset \mathbb{R}^m$. For each $p \in P$ let $x^*(p)$ be the interior maximizer of $f(\cdot; p)$. The the first order conditions

$$\begin{aligned}
\frac{\partial}{\partial x_1} f\left(x_1^*(p_1, \dots, p_m), \dots, x_n^*(p_1, \dots, p_m);\, p_1, \dots, p_m\right) &= 0 \\
&\vdots \\
\frac{\partial}{\partial x_i} f(x_1^*(p_1, \dots, p_m), \dots, x_n^*(p_1, \dots, p_m);\, p_1, \dots, p_m) &= 0 \\
&\vdots \\
\frac{\partial}{\partial x_n} f(x_1^*(p_1, \dots, p_m), \dots, x_n^*(p_1, \dots, p_m);\, p_1, \dots, p_m) &= 0
\end{aligned}$$

hold for each p. As such the left-hand side is a constant (zero) function of p, and so its partial derivatives are all zero. Differentiating the left-hand side of the ith first order condition with respect to p_k thus gives

$$\sum_{j=1}^{n} \left[\frac{\partial^2 f(x^*(p);\, p)}{\partial x_i \partial x_j} \frac{\partial x_j^*(p)}{\partial p_k} \right] + \frac{\partial^2 f(x^*(p);\, p)}{\partial x_i \partial p_k} = 0,$$

for all $i = 1, \dots, n$ and $k = 1, \dots, m$.

In matrix terms this becomes

$$\begin{bmatrix} \frac{\partial^2 f(x^*(p);p)}{\partial x_1^2} & \cdot & \frac{\partial^2 f(x^*(p);p)}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f(x^*(p);p)}{\partial x_n \partial x_1} & \cdot & \frac{\partial^2 f(x^*(p);p)}{\partial x_n^2} \end{bmatrix} \begin{bmatrix} \frac{\partial x_1^*(p)}{\partial p_1} & \cdot & \frac{\partial x_1^*(p)}{\partial p_m} \\ \vdots & & \vdots \\ \frac{\partial x_n^*(p)}{\partial p_1} & \cdot & \frac{\partial x_n^*(p)}{\partial p_m} \end{bmatrix} = - \begin{bmatrix} \frac{\partial^2 f(x^*(p);p)}{\partial x_1 \partial p_1} & \cdot & \frac{\partial^2 f(x^*(p);p)}{\partial x_1 \partial p_m} \\ \vdots & & \vdots \\ \frac{\partial^2 f(x^*(p);p)}{\partial x_n \partial p_1} & \cdot & \frac{\partial^2 f(x^*(p);p)}{\partial x_n \partial p_m} \end{bmatrix}$$

Now if the leftmost matrix has an inverse, then we may write

$$\begin{bmatrix} \frac{\partial x_1^*(p)}{\partial p_1} & \cdot & \frac{\partial x_1^*(p)}{\partial p_m} \\ \vdots & & \vdots \\ \frac{\partial x_n^*(p)}{\partial p_1} & \cdot & \frac{\partial x_n^*(p)}{\partial p_m} \end{bmatrix} = - \begin{bmatrix} \frac{\partial^2 f(x^*(p);p)}{\partial x_1^2} & \cdot & \frac{\partial^2 f(x^*(p);p)}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f(x^*(p);p)}{\partial x_n \partial x_1} & \cdot & \frac{\partial^2 f(x^*(p);p)}{\partial x_n^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial^2 f(x^*(p);p)}{\partial x_1 \partial p_1} & \cdot & \frac{\partial^2 f(x^*(p);p)}{\partial x_1 \partial p_m} \\ \vdots & & \vdots \\ \frac{\partial^2 f(x^*(p);p)}{\partial x_n \partial p_1} & \cdot & \frac{\partial^2 f(x^*(p);p)}{\partial x_n \partial p_m} \end{bmatrix}$$

to solve for all the comparative statics results.

When does this inverse exist? By the second order conditions for a maximum the matrix

$$\left[\frac{\partial^2 f(x^*;p)}{\partial x_i \partial x_j} \right] \quad \text{must be negative semidefinite.}$$

This means that it is invertible precisely when it is negative definite. Thus, we can solve for all the comparative statics results whenever the strong second order condition (definiteness) holds.

3.3 The envelope theorem

There is one more incredibly useful theorem, called the Envelope Theorem. I'll start by explaining why it's called the envelope theorem.

Given a one-dimensional parameterized family of curves,

$$f_\alpha : [0,1] \to \mathbb{R} \quad \text{where } \alpha \text{ runs over some interval } I,$$

a curve

$$h : [0,1] \to \mathbb{R}$$

is the **envelope** of the family if

i. each point on the curve h is tangent to one of the curves f_α and

ii. each curve f_α is tangent to h.

That is, for each α, there is some t and also for each t, there is some α, satisfying $f_\alpha(t) = h(t)$ and $f'_\alpha(t) = h'(t)$. If the correspondence between curves and points on the envelope is one-to-one, then we may regard h as a function of α. However, once we regard h as a function of α rather than t, the tangency condition has to be rewritten. This observation is the celebrated "Wong–Viner" theorem.

Now let $f : X \times P \to \mathbb{R}$, where X and P are real intervals, and consider the problem

$$\max_{x \in X} f(x, p).$$

We may call x the *decision variable* and p the *parameter*, or we may call x the *control* and p the *state*, or we may say that x is *endogenous* and p is *exogenous.* The function f is called the *objective function.*

Let $x^*(p)$ be an interior solution to this maximization problem. Note that it depends on the parameter p. Define

$$V(p) = f(x^*(p), p)$$

for each p. The function V is called the **optimal value function**.

For fixed x, the graph of the function $\phi_x : P \to \mathbb{R}$ via

$$\phi_x(p) = f(x, p)$$

defines a curve (or in higher dimensions of P, a surface).

The value function $V(p)$ satisfies

$$V(p) = f(x^*(p), p) = \max_x \phi_x(p).$$

The Envelope Theorem states that under appropriate conditions, the graph of the value function V will be the envelope of the family of curves

$$\{\phi_x : x \in \text{range } x^*\}.$$

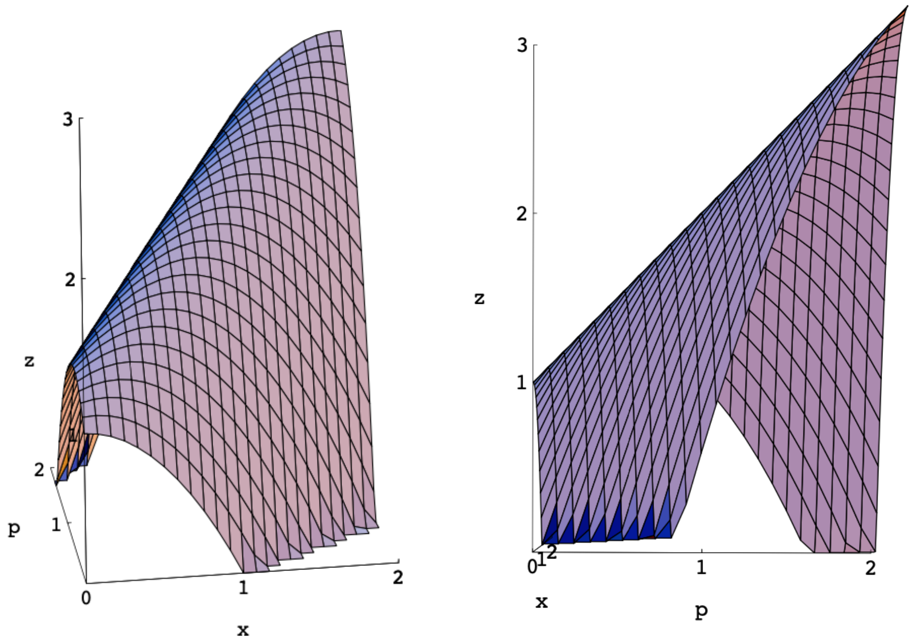
Envelope theorems in maximization theory are concerned with the tangency conditions this entails.

To get a picture of this result, imagine a plot of the graph of f. It is the surface $z = f(x, p)$ in (x, p, z)-space. Orient the graph so that the x-axis is perpendicular to the page and the p-axis runs horizontally across the page, and the z-axis is vertical. The high points of the surface (minus perspective effects) determine the graph of the value function V. Here is an example:

Example 3.4. Let

$$f(x, p) = p - (x - p)^2 + 1, \quad 0 \leq x, p \leq 2.$$

See figure 3.1a. Then given p, the maximizing x is given by $x^*(p) = p$, and $V(p) = p + 1$.



(a) Graph of $f(x, p) = p - (x - p)^2 + 1$ (b) Graph of $f(x, p) = p - (x - p)^2 + 1$ viewed from the side

Figure 3.1: Envelope graphs

The side-view of this graph in figure 3.1b shows that the high points do indeed lie on the line $z = 1 + p$. For each x, the function ϕ_x is given by

$$\phi_x(p) = p - (x - p)^2 + 1.$$

The graphs of these functions and of V are shown for selected values of x in figure 3.2. Note that the graph of V is the envelope of the family of graphs ϕ_x. Moreover the slope of V is given by

$$V'(p) = \left.\frac{\partial f}{\partial p}\right|_{x=x^*(p)=p} = 1 + 2(x - p)|_{x=p} = 1.$$

This last observation is known as the Envelope Theorem.

Theorem 3.5 (Envelope Theorem version 1)**.** *Assume that f and x^* are differentiable. Then*

$$V'(p) = D_2 f(x^*(p), p).$$

That is, the derivative of the optimal value function is simply the partial derivative of the objective function, evaluated at the optimal decision.

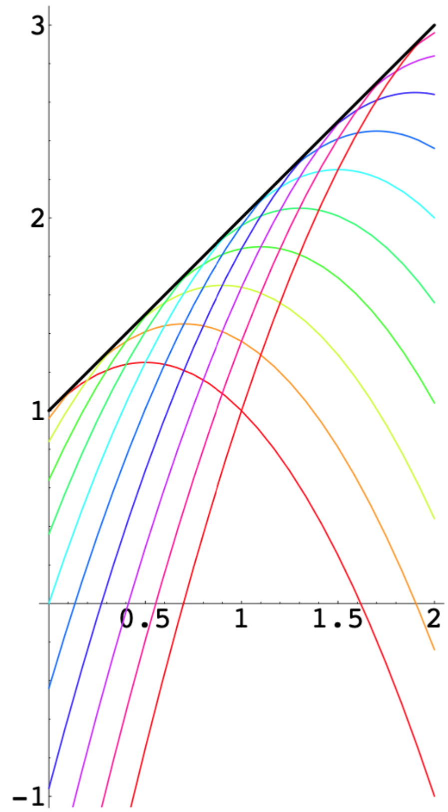


Figure 3.2: Graph of $V(p) = p + 1$ as the envelope of the family $\{\phi_x(p) : x = 0, .25, \dots, 2\}$, where $\phi_x(p) = p - (x - p)^2 + 1 = f(x, p)$.

Proof. By hypothesis f and x^* are differentiable, so V is also differentiable. By the chain rule

$$V'(p) = D_1 f(x^*(p), p) \cdot x^{*\prime}(p) + D_2 f(x^*(p), p),$$

but $D_1 f(x^*(p), p) = 0$ by the necessary first order conditions. □

Alternate proof for the one-dimensional case. By hypothesis f and x^* are differentiable, so V is also differentiable. Fix p_0 in the interior of P, and fix $x_0 = x^*(p_0)$. By definition of V, for any p,

$$V(p) \geq f(x_0, p) \text{ and } V(p_0) = f(x_0, p_0).$$

Therefore the function $h(p)$ defined by

$$h(p) = V(p) - f(x_0, p)$$

achieves its minimum at p_0. The necessary first order condition for this is that

$$V'(p_0) - D_2 f(x_0, p_0) = 0.$$

□

Note that the alternative proof generalizes easily to higher dimensional sets P, and it does not use the first-order conditions. In fact, as long as V is differentiable and f is differentiable with respect to p, the argument goes through. The key is to showing that V is differentiable.

3.4 The Envelope Theorem and the Le Chatelier Principle

Consider a producer that produces output with capital K and labor L according to the production function f.

$$\max_{K,L} pf(K, L) - wL - rK$$

Let $\pi^*(p, w, r)$ be the optimal profit function, and $K^*(p, w, r)$ and $L^*(p, w, r)$ be the **input demand functions**, and $y^*(p, w, r)$ be the **supply function**.

Now suppose K is fixed at $\bar{K} = K^*(\bar{p}, \bar{w}, \bar{r})$, and the producer wants to maximize

$$\max_{L} pf(\bar{K}, L) - wL - r\bar{K}.$$

Let $\hat{L}(p, w, r, \bar{K})$ be the **short-run demand** for labor, $\hat{y}(p, w, r, \bar{K})$ be the **short-run supply**, and $\hat{\pi}$ be the **short-run profit** function. What can we say about

$$\frac{\partial L^*}{\partial p} \text{ vs.} \frac{\partial \hat{L}}{\partial p}, \text{ etc.?}$$

Fix $\bar{w}$ and $\bar{r}$. Then

$$\pi^*(p, \bar{w}, \bar{r}) \geq \hat{\pi}(p, \bar{w}, \bar{r}, \bar{K})$$

with equality at $p = \bar{p}$. That is, $\bar{p}$ minimizes the difference, so

$$\frac{\partial \pi^*(\bar{p}, \bar{w}, \bar{r})}{\partial p} - \frac{\partial \hat{\pi}(\bar{p}, \bar{w}, \bar{r}, \bar{K})}{\partial p} = 0,$$

and

$$\frac{\partial^2 \pi^*(\bar{p}, \bar{w}, \bar{r})}{\partial p^2} - \frac{\partial^2 \hat{\pi}(\bar{p}, \bar{w}, \bar{r}, \bar{K})}{\partial p^2} \geq 0.$$

But by the Envelope Theorem,

$$\partial \pi^* / \partial p = \partial (pf(K, L) - wL - rK)/\partial p = f(K^*, L^*) = y^*,$$

so

$$\frac{\partial^2 \pi^*(\bar{p}, \bar{w}, \bar{r})}{\partial p^2} = \frac{\partial y^*(\bar{p}, \bar{w}, \bar{r})}{\partial p},$$

and

$$\frac{\partial^2 \hat{\pi}(\bar{p}, \bar{w}, \bar{r}, \bar{K})}{\partial p^2} = \frac{\partial \hat{y}(\bar{p}, \bar{w}, \bar{r}, \bar{K})}{\partial p}.$$

Thus

$$\frac{\partial y^*(\bar{p}, \bar{w}, \bar{r})}{\partial p} \geq \frac{\partial \hat{y}(\bar{p}, \bar{w}, \bar{r}, \bar{K})}{\partial p}.$$

This sort of result is known as **Le Chatelier's Principle**. Similarly, we can prove

$$\frac{\partial L^*(\bar{p}, \bar{w}, \bar{r})}{\partial w} \leq \frac{\partial \hat{L}(\bar{p}, \bar{w}, \bar{r}, \bar{K})}{\partial w}.$$

3.5 Le Chatelier without the Envelope Theorem

Long Run

$$\max \; R(L, K) - wL - rK$$

FOC

$$R_L - w = 0$$
$$R_K - r = 0$$

SOC

$$\begin{bmatrix} R_{LL} & R_{LK} \\ R_{KL} & R_{KK} \end{bmatrix} \quad \text{is negative semidefinite.}$$

Comparative statics: Differentiate the first order conditions wrt w. Write $L(w)$, $K(w)$.

$$R_{LL}L' + R_{LK}K' - 1 = 0$$
$$R_{KL}L' + R_{KK}K' = 0$$

ADD something

$$\begin{bmatrix} R_{LL} & R_{LK} \\ R_{KL} & R_{KK} \end{bmatrix} \begin{bmatrix} L' \\ K' \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} L' \\ K' \end{bmatrix} = \begin{bmatrix} R_{LL} & R_{LK} \\ R_{KL} & R_{KK} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} L' \\ K' \end{bmatrix} = \frac{1}{D} \begin{bmatrix} R_{KK} & -R_{LK} \\ -R_{KL} & R_{LL} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

So

$$L' = \frac{R_{KK}}{D} = \frac{R_{KK}}{R_{KK}R_{LL} - R_{KL}^2}$$

By SOC (and existence of inverse) we have

$$R_{KK} < 0, \quad R_{LL} < 0, \quad D > 0.$$

So

$$L'(w) < 0,$$

which is not surprising. Also

$$K'(w) = \frac{-R_{KL}}{D}$$

This sign is harder to figure. It is the opposite of R_{KL}. If w increases, the first order conditions require that the MPL increase. This is accomplished by decreasing L. This in turn changes the MPK by R_{KL}. If $R_{KL} > 0$, then a decrease in L will decrease R_K, so it is now less than r, so K must decrease to raise the MPK up to r.

Short Run

In the short run K is fixed, so the FOC is

$$R_{LL}L'_{\text{SR}} - 1 = 0$$

$$L'_{\text{SR}} = \frac{1}{R_{LL}}$$

Comparison

How do we compare this to the long run?

In the long run,

$$L' = \frac{R_{KK}}{R_{KK}R_{LL} - R_{KL}^2} = \frac{1}{R_{LL} - \frac{R_{KL}^2}{R_{KK}}} = \frac{1}{R_{LL} + \varepsilon} < 0$$

where $\varepsilon > 0$. Thus

$$0 > L' > L'_{\text{SR}}.$$

That is, the short run response of L to a change in w is greater in magnitude than the long run response.

3.6 Quasiconcave functions

There are weaker notions of convexity that are commonly applied in economic theory.

Definition 3.6. A function $f : C \to \mathbb{R}$ on a convex subset C of a vector space is:

- **quasiconcave** if for all x, y in C with $x \neq y$ and all $0 < \lambda < 1$

$$f(\lambda x + (1-\lambda)y) \geq \min\{f(x), f(y)\}.$$

- **strictly quasiconcave** if for all x, y in C with $x \neq y$ and all $0 < \lambda < 1$

$$f(\lambda x + (1-\lambda)y) > \min\{f(x), f(y)\}.$$

- **explicitly quasiconcave** or **semistrictly quasiconcave** if it is quasiconcave and in addition, for all x, y in C with $x \neq y$ and all $0 < \lambda < 1$

$$f(x) > f(y) \Rightarrow f(\lambda x + (1-\lambda)y) > \min\{f(x), f(y)\} = f(y).$$

- **quasiconvex** if for all x, y in C with $x \neq y$ and all $0 < \lambda < 1$

$$f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}.$$

- **strictly quasiconvex** if for all x, y in C with $x \neq y$ and all $0 < \lambda < 1$

$$f(\lambda x + (1-\lambda)y) < \max\{f(x), f(y)\}.$$

- **explicitly quasiconvex** or **semistrictly quasiconvex** if it is quasiconvex and in addition, for all x, y in C with $x \neq y$ and all $0 < \lambda < 1$

$$f(x) < f(y) \Rightarrow f(\lambda x + (1-\lambda)y) < \max\{f(x), f(y)\} = f(y).$$

There are other choices we could have made for the definition based on the next lemma.

Lemma 3.7. *For a function $f : C \to \mathbb{R}$ on a convex set, the following are equivalent:*

1. *The function f is quasiconcave.*
2. *For each $\alpha \in \mathbb{R}$, the strict upper contour set $[f(x) > \alpha]$ is convex, but possibly empty.*
3. *For each $\alpha \in \mathbb{R}$, the upper contour set $[f(x) \geq \alpha]$ is convex, but possibly empty.*

Proof. (1) $\Rightarrow$ (2) If f is quasiconcave and x, y in C satisfy $f(x) > \alpha$ and $f(y) > \alpha$, then for each $0 \leq \lambda \leq 1$ we have

$$f(\lambda x + (1-\lambda)y) \geq \min\{f(x), f(y)\} > \alpha.$$

(2) $\Rightarrow$ (3) Note that

$$[f \geq \alpha] = \bigcap_{n=1}^{\infty} [f > \alpha - \tfrac{1}{n}],$$

and recall that the intersection of convex sets is convex.

(3) $\Rightarrow$ (1) If $[f \geq \alpha]$ is convex for each $\alpha \in \mathbb{R}$, then for $y, z \in C$ put $\alpha = \min\{f(y), f(z)\}$ and note that $f(\lambda y + (1-\lambda)z)$ belongs to $[f \geq \alpha]$ for each $0 \leq \lambda \leq 1$. □

Corollary 3.8. A concave function is quasiconcave. A convex function is quasiconvex.

Lemma 3.9. *A strictly quasiconcave function is also explicitly quasiconcave. Likewise a strictly quasiconvex function is also explicitly quasiconvex.*

Of course, not every quasiconcave function is concave.

Example 3.10 (Explicit quasiconcavity)**.** This example sheds some light on the definition of explicit quasiconcavity. Define $f : \mathbb{R} \to [0, 1]$ by

$$f(x) = \begin{cases} 0 & x = 0 \\ 1 & x \neq 0. \end{cases}$$

If $f(x) > f(y)$, then $f(\lambda x + (1-\lambda)y) > f(y)$ for every $\lambda \in (0, 1)$ (since $f(x) > f(y)$ implies $y = 0$). But f is not quasiconcave, as $\{x : f(x) \geq 1\}$ is not convex.

Exercise 3.11. Let C be a convex set in $\mathbb{R}^m$. Let f be a lower semicontinuous quasiconcave function on C that has no local maxima. Then f is explicitly quasiconcave.

Corollary 3.12. Suppose f is concave on a convex neighborhood $C \subset \mathbb{R}^n$ of x^*, and differentiable at x^*. If $f'(x^*) = 0$, then f has a global maximum over C at x^*.

Theorem 3.13 (Local maxima of explicitly quasiconcave functions)**.** *Let $f : C \to \mathbb{R}$ be an explicitly quasiconcave function (C convex). If x^* is a local maximizer of f, then it is a global maximizer of f over C.*

Proof. Let x belong to C and suppose $f(x) > f(x^*)$. Then by the definition of explicit quasiconcavity, for any $1 > \lambda > 0$, $f(\lambda x + (1-\lambda)x^*) > f(x^*)$. Since $\lambda x + (1-\lambda)x^* \to x^*$ as $\lambda \to 0$ this contradicts the fact that f has a local maximum at x^*. $\square$

Lecture 4

Production and Returns to Scale

4.1 Production

We now start to worry about where supply comes from.

We start with the special case of a producer that *produces exactly one output from m inputs.*

Production functions

When there is only one output, a **production function** f is often used to describe feasibility. With a production function the inputs as well as the outputs are represented by nonnegative numbers. If $(x_1, \ldots, x_m)$ represent the levels of inputs $1, \ldots, m$, then $f(x_1, \ldots, x_m)$ is the quantity of output generated.

The partial derivative $D_i f(x) = \frac{\partial f(x)}{\partial x_i}$ is the **marginal product** of factor i.

An **isoquant** is just a level curve of the production function f. That is, it is a set of the form

$$\{x \in \mathbb{R}^m : f(x) = y\},$$

where y is the level of output. If the production function is **monotonic** and differentiable, then isoquants are surfaces, and we can compute their slope as follows:

For simplicity consider only two inputs, x_1 and x_2. An isoquant implicitly defines x_2 as a function of x_1 via the relation

$$f(x_1, x_2) = y.$$

Let $\hat{x}_2(x_1)$ make this explicit, that is,

$$f(x_1, \hat{x}_2(x_1)) = y \text{ for all } x_1.$$

The left hand side is now just a function of x_1, and it is a constant function. Therefore its derivative is zero. By the chain rule, then

$$D_1 f + D_2 f \cdot \hat{x}_2' = 0,$$

so

$$\hat{x}_2'(x_1) = -\frac{D_1 f}{D_2 f}\Big|_{\left(x_1, \hat{x}_2(x_1)\right)}.$$

This is thus the slope of the isoquant. It is also called the **technical rate of substitution**.

Definition 4.1 (Constant returns to scale)**.** Function f satisfies constant returns to scale if for all $x \in \mathbb{R}^m_+$ and all $\lambda > 0$,

$$f(\lambda x) = \lambda f(x).$$

Production sets

We now consider a way to describe producers that can potentially produce many commodities.

If there are m commodities, a point y in $\mathbb{R}^m$ can be used to represent a **production plan**, where y_i indicates the quantity of good i used or produced. The **sign convention** is that $y_i > 0$ indicates that good i is an output and $y_i < 0$ indicates that it is an input. The **technology set** $Y \subset \mathbb{R}^m$ is the set of **feasible** plans.

$$\begin{aligned} x \geqq y &\quad\Leftrightarrow\quad x_i \geq y_i,\ i = 1, \dots, n \\ x > y &\quad\Leftrightarrow\quad x_i \geq y_i,\ i = 1, \dots, n \text{ and } x \neq y \\ x \gg y &\quad\Leftrightarrow\quad x_i > y_i,\ i = 1, \dots, n \end{aligned}$$

Orderings on $\mathbb{R}^n$.

A plan $y \in Y$ is **(technologically) efficient** if there is no $y' \in Y$ such that $y' > y$. A **transformation function** is a function $T : \mathbb{R}^m \to \mathbb{R}$ such that $T(y) \geq 0$ if and only if $y \in Y$ and $T(y) = 0$ if and only if y is efficient.

4.2 Profit Maximization

Now introduce price vectors $p \in \mathbb{R}^m_{++}$. Describe the geometry of the dot product and how it relates to

$$\max p \cdot y \text{ over } Y.$$

The **optimal profit function** $\pi(p)$. Mention the support function theorem.

Production function approach

Introduce the **wage vector**.

$$\max_x pf(x) - w \cdot x.$$

Let x^* be the optimal input combination, known as the **factor demand function**. The **optimal profit function**

$$\pi(p, w) = pf(x^*(p, w)) - w \cdot x^*(p, w).$$

By the Envelope Theorem we have
$$\frac{\partial \pi}{\partial w_i} = -x_i^*.$$

4.3 Constant returns to scale

Recall that a production function $f : \mathbb{R}^n_+ \to \mathbb{R}$ exhibits **constant returns to scale** if for all $x \in \mathbb{R}^n_+$ and all $\lambda > 0$,

$$f(\lambda x) = \lambda f(x).$$

Letting $x = 0$ we see that $f(0) = \lambda f(0)$, so $f(0) = 0$.

Homogeneous functions

Let $f : \mathbb{R}^n_+ \to \mathbb{R}$. We say that f is **homogeneous of degree** k if for all $x \in \mathbb{R}^n_+$ and all $\lambda > 0$,

$$f(\lambda x) = \lambda^k f(x).$$

Remark 4.2 (Euler's theorem)**.** Let $f : \mathbb{R}^n_+ \to \mathbb{R}$ be continuous, and also differentiable on

$\mathbb{R}^n_{++}$. Then f is homogeneous of degree k if and only if for all $x \in \mathbb{R}^n_{++}$,

$$kf(x) = \sum_{i=1}^{n} D_i f(x) x_i. \tag{(*)}$$

Corollary 4.3. Let $f : \mathbb{R}^n_+ \to \mathbb{R}$ be continuous and differentiable on $\mathbb{R}^n_{++}$. If f is homogeneous of degree k, then $D_j f(x)$ is homogeneous of degree $k-1$.

Proposition 4.4 (Everything is constant returns to scale)**.** *Given* $f : \mathbb{R}^m \to \mathbb{R}$ *define* $g : \mathbb{R}^{\mathrm{m+1}} \to \mathbb{R}$ *by*

$$g(x_1, \ldots, x_m, z) = zf\left(\frac{x_1}{z}, \ldots, \frac{x_m}{z}\right).$$

Then

$$g(\lambda(x,z)) = \lambda z f\left(\frac{\lambda x_1}{\lambda z}, \ldots, \frac{\lambda x_m}{\lambda z}\right) = \lambda\left\{zf\left(\frac{x_1}{z}, \ldots, \frac{x_m}{z}\right)\right\} = \lambda g(x,z).$$

Quasiconcavity and constant returns to scale

The next result has applications to production functions.

Theorem 4.5. *Let* $f : \mathbb{R}^n_+ \to \mathbb{R}_+$ *be nonnegative, nondecreasing, quasiconcave, and positively homogeneous of degree* k *where* $0 < k \le 1$*. Then* f *is concave.*

Proof. Let $x, y \in \mathbb{R}^n$ and suppose first that $f(x) = \alpha > 0$ and $f(y) = \beta > 0$. (The case $\alpha = 0$ and/or $\beta = 0$ will be considered in a moment.) Then by homogeneity,

$$f\left(\frac{x}{\alpha^{\frac{1}{k}}}\right) = f\left(\frac{y}{\beta^{\frac{1}{k}}}\right) = 1$$

By quasiconcavity,

$$f\left(\lambda\frac{x}{\alpha^{\frac{1}{k}}} + (1-\lambda)\frac{y}{\beta^{\frac{1}{k}}}\right) \ge 1$$

for $0 \le \lambda \le 1$. So setting $\lambda = \frac{\alpha^{\frac{1}{k}}}{\alpha^{\frac{1}{k}} + \beta^{\frac{1}{k}}}$, we have

$$f\left(\frac{x}{\alpha^{\frac{1}{k}} + \beta^{\frac{1}{k}}} + \frac{y}{\alpha^{\frac{1}{k}} + \beta^{\frac{1}{k}}}\right) \ge 1.$$

By homogeneity,

$$f(x+y) \ge (\alpha^{\frac{1}{k}} + \beta^{\frac{1}{k}})^k = \left[f(x)^{\frac{1}{k}} + f(y)^{\frac{1}{k}}\right]^k. \tag{4.1}$$

Observe that since f is nonnegative and nondecreasing, (4.1) holds even if $f(x) = 0$ or $f(y) = 0$. Now replace x by μx and y by $(1-\mu)y$ in (4.1), where $0 \le \mu \le 1$, to get

$$\begin{aligned} f(\mu x + (1-\mu)y) &\ge \left[f(\mu x)^{\frac{1}{k}} + f((1-\mu)y)^{\frac{1}{k}}\right]^k \\ &= \left[\mu f(x)^{\frac{1}{k}} + (1-\mu) f(y)^{\frac{1}{k}}\right]^k \\ &\ge \mu \left(f(x)^{\frac{1}{k}}\right)^k + (1-\mu)\left(f(y)^{\frac{1}{k}}\right)^k \\ &= \mu f(x) + (1-\mu) f(y), \end{aligned}$$

where the last inequality follows from the concavity of $\gamma \mapsto \gamma^k$. Since x and y are arbitrary, f is concave. □

An application to the Cobb–Douglas function

Proposition 4.6. *The Cobb–Douglas function defined by*

$$f(x_1, \ldots, x_n) = x_1^{\alpha_1} x_2^{\alpha_2} \ldots, x_n^{\alpha_n},$$

where $\alpha_i > 0$, $i = 1, \ldots, n$, and $\sum_i \alpha_i \le 1$, is a concave function.

Proof. Start by observing the extended-real valued function $x \mapsto \ln x$ is strictly concave on $\mathbb{R}_+$, since its second derivative is everywhere strictly negative. Therefore the function $(x_1, \ldots, x_n) \mapsto \ln x_i$ is concave on $\mathbb{R}_+^n$ for each i. Since nonnegative scalar multiples and sums of concave functions are concave, the function

$$\phi : (x_1, \ldots, x_n) \mapsto \sum_{i=1}^{n} \alpha_i \ln x_i$$

is concave and therefore quasiconcave. Now the function $y \mapsto e^y$ is strictly monotonic, so its composition with ϕ, namely

$$f(x_1, \ldots, x_n) = e^{\phi(x)} = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n},$$

is quasiconcave. But f is homogeneous of degree $\alpha = \alpha_1 + \ldots + \alpha_n \le 1$, so by Theorem 4.5, it is concave. □

4.4 Constant Returns and Profit Maximization

An important and somewhat counterintuitive property of constant returns to scale production is this.

> If a production function f exhibits constant returns to scale and if the problem
>
> $$\max_x \pi(x) = pf(x) - w \cdot x$$
>
> has a solution, then the optimal profit is zero.

The proof is simple. By constant returns $f(0) = f(0x) = 0f(x) = 0$ for any x, so $f(0) = 0$, and it thus always possible to earn a profit of zero by setting $x = 0$. On the other hand if $pf(\bar{x}) - w \cdot \bar{x} > 0$, then no profit maximizer can exist, because if $\pi(\bar{x}) > 0$, then $\pi(2\bar{x}) = 2\pi(\bar{x}) > \pi(\bar{x})$.

So the only way a profit maximizer can exist is if the maximal profit is zero. This implies a very special relationship between p and $w_1, \dots, w_n$ must exist. For instance, in the case of one input ($n = 1$), constant returns to scale and monotonicity imply that $f(x)$ is of the form $f(x) = \alpha x$ for some $\alpha > 0$. Then the producer wants to maximize

$$p\alpha x - wx = (\alpha p - w)x$$

over the interval $[0, \infty)$. This is a linear function of x and achieves a unique maximum at $x = 0$ if $\alpha p < w$, and if $\alpha p = w$, then every $x \geq 0$ maximizes profit. In this case, the supply curve is vertical at $p = w/\alpha$, so it isn't really a supply *function*. Instead we call it a supply correspondence, and it is undefined for $p > \alpha/w$.

More generally, if there are $n > 1$ inputs and $x^* \gg 0$ maximizes profit, the first order conditions tell us that

$$p\frac{\partial f(x^*)}{\partial x_i} = w_i, \quad i = 1, \dots, n,$$

so and Euler's theorem yields

$$pf(x^*) = p\sum_{i=1}^{n} \frac{\partial f(x^*)}{\partial x_i} x_i^* = \sum_{i=1}^{n} w_i x_i^* = w \cdot x^*,$$

again the profit $pf(x^*) - w \cdot x^* = 0$.

4.5 Per capita analysis and macroeconomics

In a simple model of the macroeconomy, there is one good, **output**, which may either be consumed or saved to become part of the **capital stock** K. Output is produced from capital and labor according to the aggregate production function F,

$$Y = F(K, L),$$

Y is the flow of real output, and K is the capital stock, and L is the flow of labor supply. If F exhibits constant return to scale,

$$F(\lambda K, \lambda L) = \lambda F(K, L),$$

Euler's theorem tells us that

$$F_K(K, L)K + F_L(K, L)L = F(K, L).$$

We may also analyze the economy in per capita terms. Define

$$y = \frac{Y}{L} \qquad k = \frac{K}{L}.$$

Then

$$y = \frac{F(K, L)}{L} = F(K/L, \underbrace{L/L}_{=1}) = f(k).$$

Savings and Population Dynamics

We now make everything a function of time t. Start by assuming a constant rate of growth of the labor supply:

$$\frac{\dot{L}(t)}{L(t)} = n \qquad \text{or} \qquad L(t) = L_0 e^{nt}$$

where the dot denotes differentiation with respect to time t, and n is an exogenous constant. If there is no depreciation of capital, and a constant fraction s of output is saved, then

$$\dot{K}(t) = sY(t).$$

We may write K in terms of k as

$$K(t) = k(t)L(t)$$

which implies

$$\dot{K} = \dot{k}L + nkL.$$

But we may also write $\dot{K}$ in terms of Y as

$$\dot{K}(t) = sY(t) = sL(t)f(k(t))$$

Combining these two expressions for $\dot{K}$ gives

$$(\dot{k} + nk)L = sLf(k)$$

or

$$\dot{k} + nk = sf(k)$$

So

$$\boxed{\dot{k}(t) = sf((k(t)) - nk(t)} \tag{4.2}$$

and

$$\dot{y}(t) = f'((k(t))\dot{k}(t).$$

Example: Cobb–Douglas Production (Solow)

For the case

$$F(K, L) = K^{\alpha}L^{1-\alpha}$$

we have

$$f(k) = k^{\alpha}$$

so (4.2) becomes

$$\dot{k} = sk^{\alpha} - nk.$$

The solution to this differential equation is

$$k(t) = \left[\left(k_0^{1-\alpha} - \frac{s}{n}\right) e^{-n(1-\alpha)t} + \frac{s}{n}\right]^{\frac{1}{1-\alpha}}$$

So

$$k(t) \to k^* = \left(\frac{s}{n}\right)^{\frac{1}{1-\alpha}}$$
$$y(t) \to y^* = \left(\frac{s}{n}\right)^{\frac{\alpha}{1-\alpha}}$$

And growth stops.

Lecture 5

Convex Analysis and Support Functions

5.1 Geometry of the Euclidean inner product

The Euclidean inner product of p and x is defined by

$$p \cdot x = \sum_{i=1}^{m} p_i x_i$$

Properties of the inner product include:

1. $p \cdot p \geq 0$ and $p \neq 0 \Rightarrow p \cdot p > 0$
2. $p \cdot x = x \cdot p$
3. $p \cdot (\alpha x + \beta y) = \alpha(p \cdot x) + \beta(p \cdot y)$
4. $|p| = (p \cdot p)^{1/2}$
5. $p \cdot x = |p|\,|x|\,\cos\theta$, where θ is the angle between p and x.

To see that

$$x \cdot y = |x|\,|y|\,\cos\theta,$$

where θ is the angle between x and y, orthogonally project y on the space spanned by x. That is, write $y = \alpha x + z$ where $z \cdot x = 0$. Thus

$$z \cdot x = (y - \alpha x) \cdot x = y \cdot x - \alpha x \cdot x = 0 \quad \Rightarrow \quad \alpha = x \cdot y / x \cdot x.$$

Referring to figure 5.1 we see that

$$\cos\theta = \alpha|x|/|y| = x \cdot y/|x|\,|y|.$$

For a nonzero $p \in \mathbb{R}^m$,

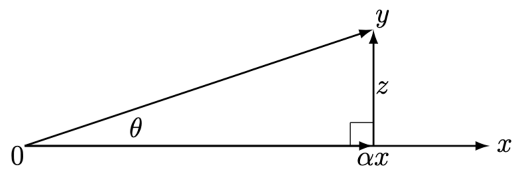


Figure 5.1: Dot product and angles

$$\{x \in \mathbb{R}^m : p \cdot x = 0\}$$

is a linear subspace of dimension $m-1$. It is the subspace of all vectors x making a right angle with p.

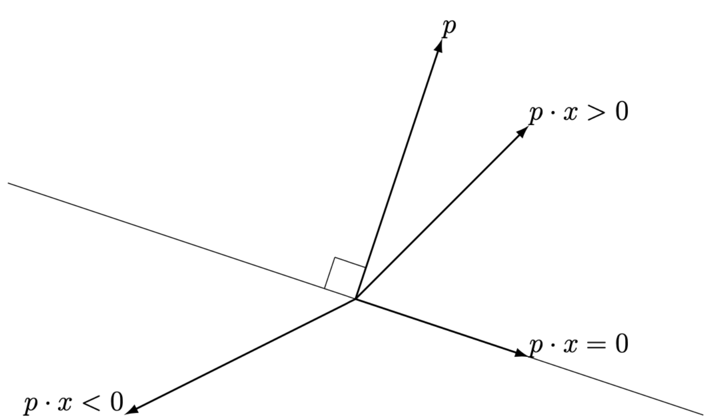


Figure 5.2: Sign of the dot product

A set of the form

$$\{x \in \mathbb{R}^m : p \cdot x = c\}, \quad p \neq 0$$

is called a **hyperplane**. To visualize the hyperplane $H = \{x : p \cdot x = c\}$ start with the vector $\alpha p \in H$, where $\alpha = c/p \cdot p$. Draw a line perpendicular to p at the point αp. For any x on this line, consider the right triangle with vertices $0, (\alpha p)$ and x. The angle x makes with p has cosine equal to $\|\alpha p\|/\|x\|$, so $p \cdot x = \|p\|\;\|x\|\;\|\alpha p\|/\|x\| = \alpha p \cdot p = c$. That is, the line lies in the hyperplane H. See figure 5.3.

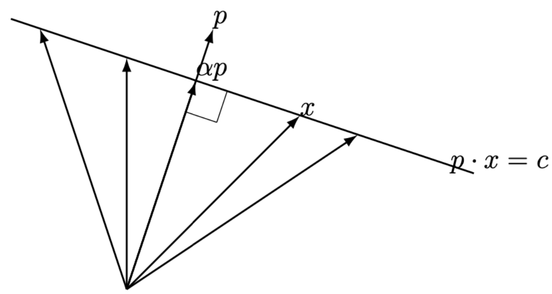


Figure 5.3: A hyperplane

5.2 Production sets

We now consider a way to describe producers that can potentially produce many commodities. Multiproduct producers are by far more common than single-product producers.

If there are m commodities altogether (inputs, outputs, intermediate goods), a point y in $\mathbb{R}^m$ can be used to represent a **production plan**, where y_i indicates the quantity of good i used or produced. The **sign convention** is that $y_i > 0$ indicates that good i is an output and $y_i < 0$ indicates that it is an input. The **technology set** $Y \subset \mathbb{R}^m$ is the set of **feasible** plans.

$$\begin{aligned} x \geqq y &\Leftrightarrow x_i \geq y_i,\ i = 1, \ldots, n \\ x > y &\Leftrightarrow x_i \geq y_i,\ i = 1, \ldots, n \text{ and } x \neq y \\ x \gg y &\Leftrightarrow x_i > y_i,\ i = 1, \ldots, n \end{aligned}$$

Orderings on $\mathbb{R}^n$.

A plan $y \in Y$ is **(technologically) efficient** if there is no $y' \in Y$ such that $y' > y$. A **transformation function** is a function $T : \mathbb{R}^m \to \mathbb{R}$ such that $T(y) \geq 0$ if and only if $y \in Y$ and $T(y) = 0$ if and only if y is efficient.

5.3 Maximizing a linear function

Given a production set $Y \subset \mathbb{R}^m$ obeying our sign convention, and a nonzero vector of prices p, the profit maximization problem is to

$$\max_{y \in Y} \ p \cdot y.$$

This is because outputs have a positive sign, so if $y_j > 0$, then $p_j y_j$ is output j's contribution to revenue, and if $y_j < 0$, then $p_j y_j$ is input j's contribution to costs. Note that for inputs, the price and wage are the same thing.

The **profit function** assigns the maximized profit to each p:

$$\pi(p) = \sup_{y \in Y} p \cdot y.$$

Geometrically it amount to finding the "highest" hyperplane orthogonal to p that touches Y. See figure 5.4.

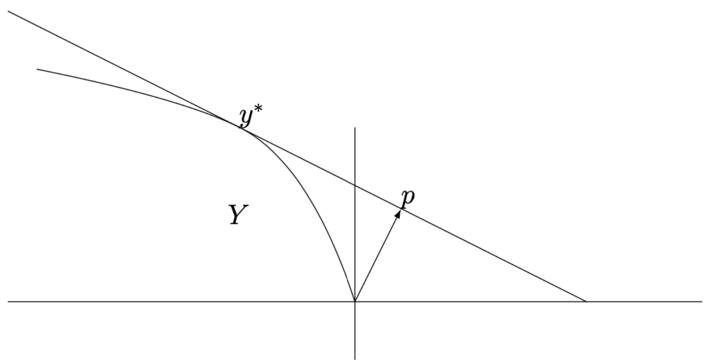


Figure 5.4: Maximizing profit

5.4 Profit and cost functions

Let A be a subset of $\mathbb{R}^m$. Convex analysts may give one of two definitions for the **support function** of A as either an infimum or a supremum. Recall that the **supremum** of a set of real numbers is its least upper bound and the **infimum** is its greatest lower bound. By convention, if A has no upper bound, $\sup A = \infty$ and if A has no lower bound, then $\inf A = -\infty$. For the empty set, $\sup A = -\infty$ and $\inf A = \infty$; otherwise $\inf A \leq \sup A$. (This makes a kind of sense: Every real number λ is an upper bound for the empty set, since there is no member of the empty set that is greater than λ. Thus the least upper bound must be $-\infty$. Similarly, every real number is also a lower bound, so the infimum is ∞.) Thus support functions (as infima or suprema) may assume the values ∞ and $-\infty$.

By convention, $0 \cdot \infty = 0$; if $\lambda > 0$ is a real number, then $\lambda \cdot \infty = \infty$ and $\lambda \cdot (-\infty) = -\infty$; and if $\lambda < 0$ is a real number, then $\lambda \cdot \infty = -\infty$ and $\lambda \cdot (-\infty) = \infty$. These conventions are used to simplify statements involving positive homogeneity.

Rather than choose one definition, I shall give the two definitions different names derived by the sort of economic interpretation I want to give them.

The **profit function** π_A of A is defined by

$$\pi_A(p) = \sup_{y \in A} p \cdot y.$$

The **cost function** c_A of A is defined by

$$c_A(p) = \inf_{y \in A} p \cdot y.$$

5.5 Introduction to convex analysis

A subset of a vector space is **convex** if it includes the line segment joining any two of its points. That is, C is convex if for each pair x, y of points in C, the **line segment**

$$\{\lambda x + (1 - \lambda)y : \lambda \in [0, 1]\}$$

is included in C. Intuitively a convex set has no holes or dents.

The **convex hull** of a set $A \subset \mathbb{R}^n$, denoted co$A$, is the smallest convex set that includes A. You can think of it as filling in any holes or dents. It consists of all points of the form

$$\sum_{i=1}^{m} \lambda_i x_i$$

where each $x_i \in A$, each $\lambda_i > 0$, and $\sum_{i=1}^{m} \lambda_i = 1$. The **closed convex hull**, $\overline{\text{co}}A$, is the smallest closed convex set that includes A. It is the **closure** of co$A$.

Result 5.1 (Carathéodory's Theorem)**.** In the above sum, m need be no larger than $n + 1$. (Remember, n is the dimension of the space.)

5.6 Concave and convex functions

An extended real-valued function f on a convex set C is **concave** if its **hypograph**

$$\{(x, \alpha) \in \mathbb{R}^m : f(x) \geq \alpha\}$$

is a convex set, or equivalently if

$$f((1-\lambda)x + \lambda y) \geq (1-\lambda)f(x) + \lambda f(y), \quad (0 < \lambda < 1).$$

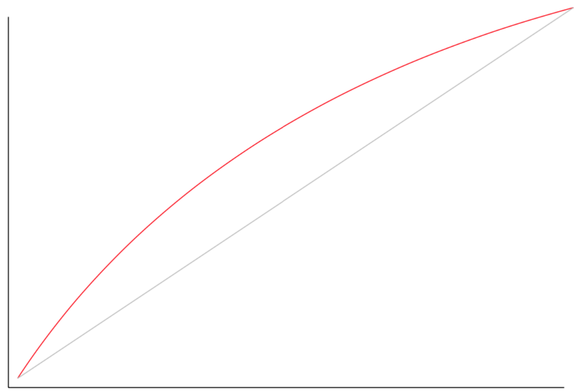


Figure 5.5: A concave function

An extended real-valued function f on a convex set C is **convex** if its **epigraph**

$$\{(x, \alpha) \in \mathbb{R}^m : f(x) \leq \alpha\}$$

is a convex set. Equivalently if

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y), \quad (0 < \lambda < 1).$$

A function f is convex if and only if $-f$ is concave. A function $f : C \to \mathbb{R}$ on a convex

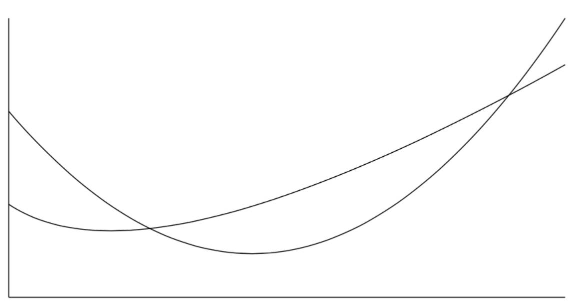


Figure 5.6: The supremum of convex functions is convex

subset C of a vector space is:

- **concave** if

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y)$$

- **strictly concave** if

$$f(\lambda x + (1-\lambda)y) > \lambda f(x) + (1-\lambda)f(y)$$

for all x, y in C with $x \neq y$ and all $0 < \lambda < 1$. It is **convex** if $-f$ is concave, etc.

Proposition 5.2. *The pointwise supremum of a family of convex functions is convex. The pointwise infimum of family of concave functions is concave.*

To see why this is true, note that the epigraph of the supremum of a family is the intersection of their epigraphs; and the intersection of convex sets is convex.

5.7 Convexity of the profit function

Proposition: π_A is convex, lower semicontinuous, and positively homogeneous of degree 1.

Proposition: c_A is concave, upper semicontinuous, and positively homogeneous of degree 1.

Proof. Fix p^0, p^1, define $p^\lambda = \lambda p^1 + (1-\lambda)p^0$, and for $\lambda \in [0,1]$, let x^λ maximize $p^\lambda \cdot x$ over A.

$$\pi_A(p^0) = p^0 \cdot x^0 \geq p^0 \cdot x^\lambda$$

$$\pi_A(p^1) = p^1 \cdot x^1 \geq p^1 \cdot x^\lambda$$

So

$$(1-\lambda)\pi_A(p^0) \geq (1-\lambda)p^0 \cdot x^\lambda$$

$$\lambda \pi_A(p^1) \geq \lambda p^1 \cdot x^\lambda$$

Adding gives:

$$\begin{aligned}
\lambda \pi_A(p^1) + (1-\lambda)\pi_A(p^0) &\geq (\lambda p^1 + (1-\lambda)p^0) \cdot x^\lambda \\
&= p^\lambda \cdot x^\lambda \\
&= \pi_A(p^\lambda) \\
&= \pi_A(\lambda p^1 + (1-\lambda)p^0)
\end{aligned}$$

□

Positive homogeneity of π_A is obvious given the conventions on multiplication of infinities. To see that it is convex, let g_x be the linear (hence convex) function defined by $g_x(p) = x \cdot p$. Then $\pi_A(p) = \sup_{x \in A} g_x(p)$. Since the pointwise supremum of a family of convex functions is convex, π_A is convex. Also each g_x is continuous, hence lower semicontinuous, and the supremum of a family of lower semicontinuous functions is lower semicontinuous.

Proposition: The set

$$\{p \in \mathbb{R}^m : \pi_A(p) < \infty\}$$

is a closed convex cone, called the **effective domain** of π_A, and denoted $\mathrm{dom}\pi_A$.

The effective domain will always include the point 0 provided A is nonempty. By convention $\pi_\emptyset(p) = -\infty$ for all p, and we say that $\pi_\emptyset$ is **improper**. If $A = \mathbb{R}^m$, then 0 is the only point in the effective domain of π_A.

It is easy to see that the effective domain $\mathrm{dom}\pi_A$ of π_A is a cone, that is, if $p \in \mathrm{dom}\pi_A$, then $\lambda p \in \mathrm{dom}\pi_A$ for every $\lambda \geq 0$. (Note that $\{0\}$ is a (degenerate) cone.)

It is also straightforward to show that $\mathrm{dom}\pi_A$ is convex. For if $\pi_A(p) < \infty$ and $\pi_A(q) < \infty$, for $0 \leq \lambda \leq 1$, by convexity of π_A, we have

$$\begin{aligned}\pi_A(\lambda p + (1-\lambda)q) &\leq \lambda \pi_A(p) + (1-\lambda)\pi_A(q) \\ &< \infty.\end{aligned}$$

Positive homogeneity of c_A is obvious given the conventions on multiplication of infinities. To see that it is concave, let g_x be the linear (hence concave) function defined by $g_x(p) = x \cdot p$. Then $c_A(p) = \inf_{x \in A} g_x(p)$. Since the pointwise infimum of a family of concave functions is concave, c_A is concave. Also each g_x is continuous, hence upper semicontinuous, and the infimum of a family of upper semicontinuous functions is upper semicontinuous.

Proposition: The set

$$\{p \in \mathbb{R}^m : c_A(p) > -\infty\}$$

is a closed convex cone, called the **effective domain** of c_A, and denoted $\mathrm{dom}c_A$.

The effective domain will always include the point 0 provided A is nonempty. By convention $c_\emptyset(p) = \infty$ for all p, and we say that $c_\emptyset$ is **improper**. If $A = \mathbb{R}^m$, then 0 is the only point in the effective domain of c_A.

It is easy to see that the effective domain $\mathrm{dom}c_A$ of c_A is a cone, that is, if $p \in \mathrm{dom}c_A$, then $\lambda p \in \mathrm{dom}c_A$ for every $\lambda \geq 0$. (Note that $\{0\}$ is a (degenerate) cone.)

It is also straightforward to show that $\mathrm{dom}c_A$ is convex. For if $c_A(p) > -\infty$ and $c_A(q) > -\infty$, for $0 \leq \lambda \leq 1$, by concavity of c_A, we have

$$\begin{aligned}c_A(\lambda x + (1-\lambda)y) &\geq \lambda c_A(p) + (1-\lambda)c_A(q) \\ &> -\infty.\end{aligned}$$

The closedness of domπ_A is more difficult.

The closedness of domc_A is more difficult.

5.8 When do maximizers and minimizers exist?

Let $K \subset \mathbb{R}^n$ be closed and bounded. Let $f : K \to \mathbb{R}$ be continuous. Then f has a maximizer and minimizer in K.

More generally, let K be a compact subset of a metric space. If f is upper semicontinuous, then f has a maximizer in K, and if f is lower semicontinuous, then f has a minimizer in K.

5.9 Recoverability

Separating Hyperplane Theorem If A is a nonempty closed convex set, and x does not belong to A, then there is a nonzero p satisfying

$$p \cdot x > \pi_A(p).$$

Proposition: The closed convex hull $\overline{co}A$ of A satisfies

$$\overline{co}A = \{y \in \mathbb{R}^m : \forall p \in \mathbb{R}^m p \cdot y \leq \pi_A(p)\}.$$

Proposition: If f is continuous on its effective domain, convex, and positively homogeneous of degree 1, define

$$A = \{y \in \mathbb{R}^m : \forall p \in \mathbb{R}^m p \cdot y \leq f(p)\}.$$

Then A is closed and convex and

$$f = \pi_A.$$

Separating Hyperplane Theorem If A is a nonempty closed convex set, and x does not belong to A, then there is a nonzero p satisfying

$$p \cdot x < c_A(p).$$

Proposition: The closed convex hull $\overline{co}A$ of A satisfies

$$\overline{co}A = \{y \in \mathbb{R}^m : \forall p \in \mathbb{R}^m p \cdot y \geq c_A(p)\}.$$

Proposition: If f is continuous on its effective domain, concave, and positively homogeneous of degree 1, define

$$A = \{y \in \mathbb{R}^m : \forall p \in \mathbb{R}^m p \cdot y \geq f(p)\}.$$

Then A is closed and convex and

$$f = c_A.$$

5.10 Recovering an input requirement set

The **input requirement set** is the set of inputs that allow the producer to produce an output level of at least y. Knowing the cost function as a function of w, the vector of input wages. figure 5.7a shows the intersection of a few sets of the form $\{x \in \mathbb{R}^m : w \cdot x \geq c(w; y)\}$ for the production function $f(x_1, x_2) = x_1^{1/2} x_2^{1/2}$. But if the input requirement set is not convex, you will recover its closed convex hull, see figure 5.7b, for the production function $f(x_1, x_2) = \max\{x_1^{5/6} x_2^{1/6}, x_1^{1/6} x_2^{5/6}\}$.

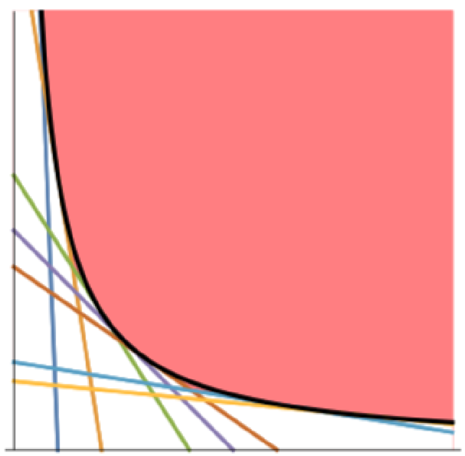


Figure 5.7.

(a)

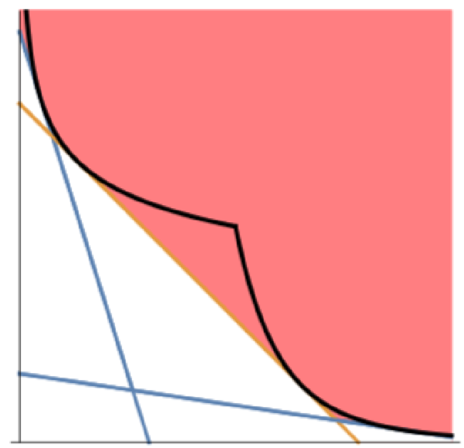


Figure 5.8.

(b)

Figure 5.7

5.11 Concavity and maxima

Proposition 5.3. *If f is a concave function on a convex set and x^* is a local maximizer, then it is a global maximizer.*

Proof. Prove the contrapositive: Suppose x^* is not a global maximizer. Let

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

Then for $0 < \lambda < 1$,

$$\lambda f(\hat{x}) + (1 - \lambda) f(x^*) > f(x^*).$$

By concavity

$$f(\lambda\hat{x} + (1-\lambda)x^*) \geq \lambda f(\hat{x}) + (1-\lambda)f(x^*) > f(x^*),$$

but $\lambda\hat{x} + (1-\lambda)x^* \to x^*$ as $\lambda \to 0$, so x^* is not a local maximizer. □

5.12 Supergradients and first order conditions

Definition 5.4. Let $f : \mathbb{R}^n \to \mathbb{R}$ be concave. A vector p is a **supergradient** of f at the point x if for every y it satisfies the **supergradient inequality**,

$$f(x) + p \cdot (y - x) \geq f(y).$$

Similarly, if f is convex, then p is a **subgradient** of f at x if

$$f(x) + p \cdot (y - x) \leq f(y)$$

for every y.

For concave f, the set of all supergradients of f at x is called the **superdifferential** of f at x, and is denoted $\partial f(x)$. If the superdifferential is nonempty at x, we say that f is **superdifferentiable** at x.

For convex f the same symbol $\partial f(x)$ denotes the set of subgradients and is called the **subdifferential**. If it is nonempty we say that f is **subdifferentiable**.

Theorem 5.5 (Gradients are supergradients)**.** *Assume f is concave on a convex set $C \subset \mathbb{R}^n$, and differentiable at the point x. Then for every y in C,*

$$f(x) + f'(x) \cdot (y - x) \geq f(y). \tag{5.1}$$

If instead f is convex, then the above inequality is reversed.

Proof. Let $y \in C$. Rewrite the definition of concavity as

$$f\big(x + \lambda(y - x)\big) \geq f(x) + \lambda\big(f(y) - f(x)\big).$$

Rearranging and dividing by $\lambda > 0$,

$$\frac{f\big(x + \lambda(y - x)\big) - f(x)}{\lambda} \geq f(y) - f(x).$$

Letting $\lambda \downarrow 0$, the left hand side converges to $f'(x) \cdot (y - x)$. □

For concave/convex functions the first order conditions for an extremum are sufficient.

Theorem 5.6 (First order conditions for concave functions)**.** *Suppose f is concave on a convex set $C \subset \mathbb{R}^n$. A point x^* in C is a global maximum point of f if and only if* 0 *belongs to the superdifferential $\partial f(x^*)$.*

Suppose f is convex on a convex set $C \subset \mathbb{R}^n$. A point x^* in C is a global minimum point of f if and only if 0 *belongs to the superdifferential $\partial f(x^*)$.*

Proof. Note that x^* is a global maximum point of f if and only if

$$f(x^*) + 0 \cdot (y - x^*) \geq f(y)$$

for all y in C, but this is just the supergradient inequality for 0. □

Corollary 5.7. If f is concave and $f'(x^*) = 0$, then x^* is a global maximizer. If f is convex and $f'(x^*) = 0$, then x^* is a global minimizer.

Proof. The graph of a concave function lies below a horizontal line at x^*, see (5.1). □

A function need not be differentiable to have sub/supergradients.

Theorem 5.8 (Subdifferentiability)**.** *A convex function on a convex subset of $\mathbb{R}^n$ is subdifferentiable at each point of the relative interior of its domain.*

A concave function on a convex subset of $\mathbb{R}^n$ is superdifferentiable at each point of the relative interior of its domain.

Fact 5.9. If f is concave, then f differentiable at x if and only if its superdifferential $\partial f(x)$ is a singleton, in which case $\partial f(x) = f'(x)$.

If f is convex, then f differentiable at x if and only if its subdifferential $\partial f(x)$ is a singleton, in which case $\partial f(x) = f'(x)$.

5.13 Jensen's Inequality

Result 5.10. Let f be a concave function, and let X be a random variable taking values in the domain of f, with $|\mathsf{E}X| < \infty$. Then

$$f(\mathsf{E}X) \geq \mathsf{E}f(X).$$

"Proof:" Evaluate (5.1) at $\mathsf{E}X$:

$$f(\mathsf{E}X) + f'(\mathsf{E}X)(X - \mathsf{E}X) \geq f(X) \quad \text{for all} X$$

and take expectations:

$$f(\mathsf{E}X) + f'(\mathsf{E}X)\underbrace{\mathsf{E}(X - \mathsf{E}X)}_{=0} \geq \mathsf{E}f(X).$$

(The result is true even f is not differentiable at $\mathsf{E}X$.) □

5.14 Concavity and second derivatives

Fact 5.11. Let f be differentiable on an open interval in $\mathbb{R}$.

Then f is concave if and only if f' is nonincreasing.

If f is concave and twice differentiable at x, then $f''(x) \leq 0$.

If f is everywhere twice differentiable with $f'' \leq 0$, then f is concave.

If f is everywhere twice differentiable with $f'' < 0$, then f is strictly concave.

Fact 5.12. If $f : C \subset \mathbb{R}^n \to \mathbb{R}$ is twice differentiable, then the Hessian H_f is everywhere negative semidefinite if and only if f is concave. If H_f is everywhere negative definite, then f is strictly concave.

5.15 The subdifferential of the profit function

$$\partial\pi_A(p) = \{x \in \overline{co}A : p \cdot x = \pi_A(p)\}$$

Extremizers are subgradients

Proposition: If $\tilde{y}(p)$ maximizes p over A, that is, if $\tilde{y}(p)$ belongs to A and $p \cdot \tilde{y}(p) \geq p \cdot y$ for all $y \in A$, then $\tilde{y}(p)$ is a subgradient of π_A at p. That is,

$$\pi_A(p) + \tilde{y}(p) \cdot (q - p) \leq \pi_A(q) \qquad (*)$$

for all $q \in \mathbb{R}^m$.

To see this, note that for any $q \in \mathbb{R}^m$, by definition we have

$$q \cdot \tilde{y}(p) \leq \pi_A(q).$$

Now add $\pi_A(p) - p \cdot \tilde{y}(p) = 0$ to the left hand side to get the subgradient inequality.

Note that $\pi_A(p)$ may be finite for a closed convex set A, and yet there may be no maximizer. For instance, let

$$A = \{(x, y) \in \mathbb{R}^2 : x < 0,\ y < 0,\ xy \geq 1\}.$$

Then for $p = (1, 0)$, we have $\pi_A(p) = 0$ as $(1, 0) \cdot (-1/n, -n) = -1/n$, but $(1, 0) \cdot (x, y) = x < 0$ for each $(x, y) \in A$. Thus there is no maximizer in A.

It turns out that if there is no maximizer of p, then π_A has no subgradient at p. In fact, the following is true, but I won't present the proof, which relies on the Separating Hyperplane Theorem.

Theorem: If A is closed and convex, then x is a subgradient of π_A at p if and only if $x \in A$ and x maximizes p over A.

Proposition: If $\hat{y}(p)$ minimizes p over A, that is, if $\hat{y}(p)$ belongs to A and $p \cdot \hat{y}(p) \leq p \cdot y$ for all $y \in A$, then $\hat{y}(p)$ is a supergradient of c_A at p. That is,

$$c_A(p) + \hat{y}(p) \cdot (q - p) \geq c_A(q) \qquad (*)$$

for all $q \in \mathbb{R}^m$.

To see this, note that for any $q \in \mathbb{R}^m$, by definition we have

$$q \cdot \hat{y}(p) \geq c_A(q).$$

Now add $c_A(p) - p \cdot \hat{y}(p) = 0$ to the left hand side to get the supergradient inequality.

Note that $x_A(p)$ may be finite for a closed convex set A, and yet there may be no minimizer. For instance, let

$$A = \{(x, y) \in \mathbb{R}^2 : x > 0,\ y > 0,\ xy \geq 1\}.$$

Then for $p = (1, 0)$, we have $\pi_A(p) = 0$ as $(1, 0) \cdot (1/n, n) = 1/n$, but $(1, 0) \cdot (x, y) = x > 0$ for each $(x, y) \in A$. Thus there is no minimizer in A.

It turns out that if there is no minimizer of p, then c_A has no supergradient at p. In fact, the following is true, but I won't present the proof, which relies on the Separating Hyperplane Theorem.

Theorem: If A is closed and convex, then x is a supergradient of c_A at p if and only if $x \in A$ and x minimizes p over A.

5.16 Hotelling's Lemma

If x is the unique profit maximizer at prices p in the convex set A, then π_A is differentiable at p and $\nabla \pi_A(p) = x$.

$$\left[D_{ij}\pi^*(p)\right] = \left[D_j y_i^*(p)\right] = \left[\tfrac{\partial y_i^*(p)}{\partial p_j}\right]$$

is symmetric and positive semidefinite. Consequently,

$$D_i y_i^*(p) = \frac{\partial y_i^*(p)}{\partial p_i} \geq 0.$$

Comparative statics

Proposition: Consequently, if A is closed and convex, and $\tilde{y}(p)$ is the unique maximizer of p over A, then π_A is differentiable at p and

$$\tilde{y}(p) = \pi_A'(p). \qquad (**)$$

Proposition: Consequently, if A is closed and convex, and $\hat{y}(p)$ is the unique minimizer of p over A, then c_A is differentiable at p and

$$\hat{y}(p) = c_A'(p). \qquad (**)$$

One way to see this is to consider q of the form $p \pm \lambda e^i$, where e^i is the i-th unit coordinate vector, and $\lambda > 0$.
The subgradient inequality for $q = p + \lambda e^i$ is

$$\tilde{y}(p) \cdot \lambda e^i \leq \pi_A(p + \lambda e^i) - \pi_A(p)$$

and for $q = p - \lambda e^i$ is

$$-\tilde{y}(p) \cdot \lambda e^i \leq \pi_A(p - \lambda e^i) - \pi_A(p).$$

Dividing these by λ and $-\lambda$ respectively yields

$$y_i^*(p) \leq \frac{\pi_A(p + \lambda e^i) - \pi_A(p)}{\lambda}$$
$$y_i^*(p) \geq \frac{\pi_A(p - \lambda e^i) - \pi_A(p)}{\lambda}.$$

so

$$\frac{\pi_A(p-\lambda e^i)-\pi_A(p)}{\lambda} \leq y_i^*(p) \leq \frac{\pi_A(p+\lambda e^i)-\pi_A(p)}{\lambda}.$$

Letting $\lambda \downarrow 0$ yields $\tilde{y}_i(p) = D_i \pi_A(p)$.

Proposition: Thus if π_A is twice differentiable at p, that is, if the maximizer $\tilde{y}(p)$ is differentiable with respect to p, then the i-th component satisfies

$$D_j y_i^*(p) = D_{ij} \pi_A(p). \qquad (***)$$

Consequently, the matrix

$$\left[D_j y_i^*(p) \right]$$

is positive semidefinite.

In particular,

$$D_i \tilde{y}_i \geq 0.$$

One way to see this is to consider q of the form $p \pm \lambda e^i$, where e^i is the i-th unit coordinate vector, and $\lambda > 0$.
The supergradient inequality for $q = p + \lambda e^i$ is

$$\hat{y}(p) \cdot \lambda e^i \geq c_A(p + \lambda e^i) - c_A(p)$$

and for $q = p - \lambda e^i$ is

$$-\hat{y}(p) \cdot \lambda e^i \geq c_A(p - \lambda e^i) - c_A(p).$$

Dividing these by λ and $-\lambda$ respectively yields

$$y_i^*(p) \geq \frac{c_A(p + \lambda e^i) - c_A(p)}{\lambda}$$
$$y_i^*(p) \leq \frac{c_A(p - \lambda e^i) - c_A(p)}{\lambda}.$$

so

$$\frac{c_A(p+\lambda e^i)-c_A(p)}{\lambda} \leq y_i^*(p) \leq \frac{c_A(p-\lambda e^i)-c_A(p)}{\lambda}.$$

Letting $\lambda \downarrow 0$ yields $\hat{y}_i(p) = D_i c_A(p)$.

Proposition: Thus if c_A is twice differentiable at p, that is, if the minimizer $\hat{y}(p)$ is differentiable with respect to p, then the i-th component satisfies

$$D_j y_i^*(p) = D_{ij} c_A(p). \qquad (***)$$

Consequently, the matrix

$$\left[D_j y_i^*(p) \right]$$

is negative semidefinite.

In particular,

$$D_i \hat{y}_i \leq 0.$$

Even without twice differentiability, from the subgradient inequality, we have

$$\pi_A(p) + \tilde{y}(p) \cdot (q - p) \leq \pi_A(q)$$
$$\pi_A(q) + \tilde{y}(q) \cdot (p - q) \leq \pi_A(p)$$

so adding the two inequalities, we get

$$\big(\tilde{y}(p) - \tilde{y}(q)\big) \cdot (p - q) \geq 0.$$

Proposition: Thus if q differs from p only in its i-th component, say $q_i = p_i + \triangle p_i$, then we have

$$\triangle\tilde{y}_i \triangle p_i \geq 0.$$

Dividing by the positive quantity $(\triangle p_i)^2$ does not change this inequality, so

$$\frac{\triangle\tilde{y}_i}{\triangle p_i} \geq 0.$$

Even without twice differentiability, from the supergradient inequality, we have

$$c_A(p) + \hat{y}(p) \cdot (q - p) \geq c_A(q)$$
$$c_A(q) + \hat{y}(q) \cdot (p - q) \geq c_A(p)$$

so adding the two inequalities, we get

$$\big(\hat{y}(p) - \hat{y}(q)\big) \cdot (p - q) \leq 0.$$

Proposition: Thus if q differs from p only in its i-th component, say $q_i = p_i + \triangle p_i$, then we have

$$\triangle\hat{y}_i \triangle p_i \leq 0.$$

Dividing by the positive quantity $(\triangle p_i)^2$ does not change this inequality, so

$$\frac{\triangle\hat{y}_i}{\triangle p_i} \leq 0.$$

Lecture 6

Production Functions, Cost Minimization, and Lagrange Multipliers

6.1 Cost minimization and convex analysis

When there is a production function f for a single output producer with n inputs, the input requirement set for producing output level y is

$$V(y) = \{x \in \mathbb{R}^n : f(x) \geq y\}.$$

The cost function for the producer facing wage vector $w = (w_1, \dots, w_n)$ is the support function

$$c(w, y) = \inf\{w \cdot x : f(x) \geq y\}.$$

The Support Function Theorem tells us that holding y fixed, c is concave in w, and if x^* is the unique cost minimizer, then

$$\frac{\partial c}{\partial w_i} = x_i^*$$

and when c is twice differentiable in w, the Hessian matrix

$$\left[\frac{\partial^2 c}{\partial w_i \partial w_j}\right] = \left[\frac{\partial x_i^*}{\partial w_j}\right] \quad \text{is symmetric and negative semidefinite.}$$

But the Support Function Theorem doesn't tell us that c is twice differentiable or how it depends on y. When the production function f is differentiable, we can use the Lagrange

Multiplier Theorem to find additional results.

6.2 Classical Lagrange Multiplier Theorem

Definition 6.1. A **constrained optimization problem** is characterized by an **objective function** f and m **constraint functions**, $g_1, \dots, g_m$. The constraints take the form of either **equality constraints** ($g_i(x) = 0$, $i = 1, \dots, m$) or **inequality constraints** ($g_i(x) \geq 0$, $i = 1, \dots, m$).

A point x^* is a **constrained local maximizer** of f subject to the equality constraints $g_1(x) = 0$, $g_2(x) = 0$, $\dots$, $g_m(x) = 0$ in some neighborhood W of x^* if x^* satisfies the constraints and also satisfies $f(x^*) \geq f(x)$ for all $x \in W$ that also satisfy the constraints.

A **constrained local minimizer** is defined similarly, and the case of inequality constraints is also dealt with as you should expect.

Frequently, the true constraints are inequality constraints, but we can see that at an extremum, those will be satisfied as equalitiesx, and we may write them as equality constraints.

Associated with such a problem is a function called the **Lagrangian**:

$$L(x; \lambda) = f(x) + \lambda \cdot g(x) = f(x) + \lambda_1 g_1(x) + \cdots + \lambda_m g_m(x).$$

The numbers λ_i are called **Lagrange multipliers**.

Result 6.2 (Lagrange Multiplier Theorem)**.** Let $X \subset \mathbb{R}^n$, and let $f, g_1, \dots, g_m : X \to \mathbb{R}$ be continuous. Let x^* be an interior constrained local maximizer of f subject to $g(x) = 0$. Suppose f, $g_1, \dots, g_m$ are differentiable at x^*, and that the Lagrange Constraint Qualification holds, that is, ${g_1}'(x^*), \dots, {g_m}'(x^*)$ are linearly independent.

Then there exist real numbers $\lambda_1^*, \dots, \lambda_m^*$, such that

$$f'(x^*) + \sum_{i=1}^{m} \lambda_i^* {g_i}'(x^*) = 0.$$

Remark 6.3. The way I wrote the Lagrangian above is the preferred way to write the Lagrangian for maximization. For minimization, the preferred way to write the Lagrangian is

$$L(x; \lambda) = f(x) - \lambda \cdot g(x) = f(x) - \lambda_1 g_1(x) - \cdots - \lambda_m g_m(x).x$$

There is no need to do this unless you care bout the sign of the multipliers (and I do). Also, the constraint $g(x) = 0$ is the same as the constraint $-g(x) = 0$, so when deciding how to write the constraint, if there is a true inequality constraint $g(x) \geq 0$ that we know a priori must hold with equality, write the equality constraint as $g(x) = 0$. This will become clearer wen you look at the examples in what follows.

6.3 Using the LMT

Since the LMT tells us what is true at the optimum, we can sometimes use the necessary conditions to pin down what the optimum is. For example, the Cobb–Douglas production function is given by

$$y = f(x) = \gamma x_1^{\alpha_1} \cdots x_n^{\alpha_n},$$

where each $\alpha_i > 0$, $i = 1, \ldots, n$. It is homogeneous of degree

$$\alpha = \sum_{i=1}^{n} \alpha_i.$$

This function was proposed by Charles Cobb and Paul Douglas as a model for U.S. GDP, and it works surprisingly well empirically. When $\gamma = \alpha = 1$, it is a weighted geometric mean of the inputs.

To find the associated cost function we start by writing the Lagrangian for a minimum, where the true constraint is $f(x) - y \geq 0$, as

$$L(x; \lambda) = w \cdot x - \lambda(\gamma x_1^{\alpha_1} \cdots x_n^{\alpha_n} - y)$$

The first order conditions, using the binding constraint $y = \gamma x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ are:

$$\frac{\partial L}{\partial x_i} = w_i - \lambda \alpha_i \frac{y}{x_i} = 0 \quad i = 1, \ldots, n.$$

So

$$x_i = \lambda \alpha_i \frac{y}{w_i} \quad i = 1, \ldots, n. \tag{6.1}$$

But $y = \gamma {x_1}^{\alpha_1} \cdots {x_n}_n^{\alpha}$, so

$$y = \gamma \prod_{i=1}^{n} \left(\lambda \alpha_i \frac{y}{w_i} \right)^{\alpha_i} = \gamma \lambda^{\alpha} y^{\alpha} \prod_{i=1}^{n} \left(\frac{\alpha_i}{w_i} \right)^{\alpha_i}.$$

Solving this for λ gives

$$\begin{aligned}\hat{\lambda} &= \left[\gamma y^{\alpha-1}\prod_{i=1}^{n}\left(\frac{\alpha_i}{w_i}\right)^{\alpha_i}\right]^{-1/\alpha}\\ &= \gamma^{-1/\alpha}y^{(1-\alpha)/\alpha}\left(\prod_{i=1}^{n}\alpha_i^{-\alpha_i/\alpha}\right)\left(\prod_{i=1}^{n}w^{\alpha_i/\alpha}\right)\end{aligned}$$

To simplify notation a bit, set

$$\begin{aligned}\beta_i &= \frac{\alpha_i}{\alpha},\\ b &= \gamma^{\frac{-1}{\alpha}}\cdot\prod_i \alpha_i^{-\beta_i},\end{aligned}$$

so

$$\hat{\lambda} = by^{(1-\alpha)/\alpha}\prod_{i=1}^{n}w_i^{\beta_i}.$$

Substituting this for λ in (6.1) gives the conditional factor demands

$$\begin{aligned}\hat{x}_j(y,w) &= by^{(1-\alpha)/\alpha}\prod_{i=1}^{n}w_i^{\beta_i}\alpha_j\frac{y}{w_j}\\ &= \frac{\alpha_j}{w_j}by^{1/\alpha}\prod_{i=1}^{n}w_i^{\beta_i},\end{aligned}$$

for $j = 1,\dots,n$. So the cost function is

$$c(y,w) = \alpha by^{1/\alpha}\prod_{i=1}^{n}w_i^{\beta_i},$$

which is a Cobb–Douglas function of ws.

Note that

$$\frac{\partial c(y,w)}{\partial y} = by^{(1-\alpha)/\alpha}\prod_{i=1}^{n}w_i^{\beta_i} = \hat{\lambda},$$

and

$$\frac{\partial c(y,w)}{\partial w_j} = \alpha\frac{\beta_j}{w_j}by^{(1-\alpha)/\alpha}\prod_{i=1}^{n}w_i^{\beta_i} = \hat{x}_j(y,w).$$

6.4 Second Order Conditions

Theorem 6.4 (Necessary Second Order Conditions for a Maximum)**.** *Let $U \subset \mathbb{R}^n$ and let $x^* \in \text{int}\, U$. Let $f, g_1, \ldots, g_m : U \to \mathbb{R}$ be C^2, and suppose x^* is a local constrained maximizer of f subject to $g(x) = 0$. Define the Lagrangian $L(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i g_i(x)$. Assume that ${g_1}'(x^*), \ldots, {g_m}'(x^*)$ are linearly independent, so the conclusion of the Lagrange Multiplier Theorem holds, that is, there are $\lambda_1^*, \ldots, \lambda_m^*$ satisfying the first order conditions*

$$L'_x(x^*, \lambda^*) = f'(x^*) + \sum_{i=1}^m \lambda_i^* {g_i}'(x^*) = 0.$$

Then

$$\sum_{i=1}^n \sum_{j=1}^n D_{ij} L(x^*, \lambda^*) v_i v_j \leq 0,$$

for all $v \neq 0$ satisfying ${g_i}'(x^*) \cdot v = 0$, $i = 1, \ldots, m$.

Since minimizing f is the same as maximizing $-f$, we do not need any new results for minimization, but there a few things worth pointing out.

The Lagrangian for maximizing $-f$ subject to $g_i = 0$, $i = 1, \ldots, m$ is

$$-f(x) + \sum_{i=1}^m \lambda_i g_i(x),$$

The second order condition for maximizing $-f$ is that

$$\sum_{i=1}^n \sum_{j=1}^n \left(-D_{ij} f(x^*) + \sum_{i=1}^m \lambda^* D_{ij} g(x^*) \right) v_i v_j \leq 0,$$

for all $v \neq 0$ satisfying ${g_i}'(x^*) \cdot v = 0$, $i = 1, \ldots, m$. This can be rewritten as

$$\sum_{i=1}^n \sum_{j=1}^n \left(D_{ij} f(x^*) - \sum_{i=1}^m \lambda^* D_{ij} g(x^*) \right) v_i v_j \geq 0,$$

which explains why I prefer to write the Lagrangian for a minimization problem as

$$L(x, \lambda) = f(x) - \sum_{i=1}^m \lambda_i g_i(x).$$

The first order conditions will be exactly the same. For the second order conditions we have the following.

Theorem 6.5 (Necessary Second Order Conditions for a Minimum). *Let $U \subset \mathbb{R}^n$ and let $x^* \in \operatorname{int} U$. Let $f, g_1, \ldots, g_m : U \to \mathbb{R}$ be C^2, and suppose x^* is a local constrained minimizer of f subject to $g(x) = 0$. Define the Lagrangian*

$$L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i g_i(x).$$

Assume that ${g_1}'(x^*), \ldots, {g_m}'(x^*)$ are linearly independent, so the conclusion of the Lagrange Multiplier Theorem holds, that is, there are $\lambda_1^*, \ldots, \lambda_m^*$ satisfying the first order conditions

$$L'_x(x^*, \lambda^*) = f'(x^*) - \sum_{i=1}^{m} \lambda_i^* {g_i}'(x^*) = 0.$$

Then

$$\sum_{i=1}^{n} \sum_{j=1}^{n} D_{ij} L(x^*, \lambda^*) v_i v_j \geq 0,$$

for all $v \neq 0$ satisfying ${g_i}'(x^*) \cdot v = 0$, $i = 1, \ldots, m$.

6.5 Envelope Theorem for Constrained Extrema

Theorem 6.6 (Envelope Theorem for Constrained Maximization). *Let $X \subset \mathbb{R}^n$ and $P \subset \mathbb{R}^l$ be open, and let $f, g_1, \ldots, g_m : X \times P \to \mathbb{R}$ be C^1. For each $p \in P$, let $x^*(p)$ be an interior constrained local maximizer of $f(x, p)$ subject to $g(x, p) = 0$. Define the Lagrangian*

$$L(x, \lambda; p) = f(x, p) + \sum_{i=1}^{m} \lambda_i g_i(x, p),$$

and assume that the conclusion of the Lagrange Multiplier Theorem holds for each p, that is, there exist real numbers $\lambda_1^*(p), \ldots, \lambda_m^*(p)$, such that the first order conditions

$$\frac{\partial L\big(x^*(p), \lambda^*(p), p\big)}{\partial x} = f'_x\big(x^*(p), p\big) + \sum_{i=1}^{m} \lambda_i^*(p) g'_{i x}\big(x^*(p), p\big) = 0$$

are satisfied. Assume that $x^* : P \to X$ and $\lambda^* : P \to \mathbb{R}^m$ are C^1. Set

$$V(p) = f\left(x^*(p), p\right).$$

Then V is C^1 and

$$\frac{\partial V(p)}{\partial p_j} = \frac{\partial L\big(x^*(p), \lambda^*(p), p\big)}{\partial p_j} = \frac{\partial f(x^*, p)}{\partial p_j} + \sum_{i=1}^{m} \lambda_i^*(p) \frac{\partial g_i(x^*, p)}{\partial p_j}.$$

Proof. Clearly V is C^1 as the composition of C^1 functions. Since x^* satisfies the constraints, we have

$$V(p) = f\left(x^*(p), p\right) = f\left(x^*(p), p\right) + \sum_{i=1}^{m} \lambda_i^*(p) g_i(x^*, p).$$

Therefore by the chain rule,

$$\begin{aligned}
\frac{\partial V(p)}{\partial p_j} &= \left(\sum_{k=1}^{n} \frac{\partial f(x^*, p)}{\partial x_k} \frac{\partial x^{*k}}{\partial p_j}\right) + \frac{\partial f(x^*, p)}{\partial p_j} \\
&\quad + \sum_{i=1}^{m} \left\{ \frac{\partial \lambda_i^*(p)}{\partial p_j} g_i(x^*, p) + \lambda^*(p) \left[\left(\sum_{k=1}^{n} \frac{\partial g_i(x^*, p)}{\partial x_k} \frac{\partial x^{*k}}{\partial p_j} \right) + \frac{\partial g_i(x^*, p)}{\partial p_j} \right] \right\} \\
&= \frac{\partial f(x^*, p)}{\partial p_j} + \sum_{i=1}^{m} \lambda_i^*(p) \frac{\partial g_i(x^*, p)}{\partial p_j} \\
&\quad + \sum_{i=1}^{m} \frac{\partial \lambda_i^*(p)}{\partial p_j} g_i(x^*, p) \qquad (6.2) \\
&\quad + \sum_{k=1}^{n} \left(\frac{\partial f(x^*, p)}{\partial x_k} + \sum_{i=1}^{m} \lambda^*(p) \frac{\partial g_i(x^*, p)}{\partial x_k} \right) \frac{\partial x^{*k}}{\partial p_j}. \qquad (6.3)
\end{aligned}$$

The theorem now follows from the fact that both terms (6.2) and (6.3) are zero. Term (6.2) is zero since each g_i is zero as x^* satisfies the constraints, and term (6.3) is zero, as the first order conditions imply that each $\frac{\partial f(x^*,p)}{\partial x_k} + \sum_{i=1}^{m} \lambda^*(p) \frac{\partial g_i(x^*,p)}{\partial x_k} = 0$. □

Theorem 6.7 (Envelope Theorem for Minimization)**.** *Let $X \subset \mathbb{R}^n$ and $P \subset \mathbb{R}^l$ be open, and let $f, g_1, \ldots, g_m : X \times P \to \mathbb{R}$ be C^1. For each $p \in P$, let $x^*(p)$ be an interior constrained local maximizer of $f(x, p)$ subject to $g(x, p) = 0$. Define the Lagrangian*

$$L(x, \lambda; p) = f(x, p) - \sum_{i=1}^{m} \lambda_i g_i(x, p),$$

and assume that the conclusion of the Lagrange Multiplier Theorem holds for each p, that is, there exist real numbers $\lambda_1^*(p), \ldots, \lambda_m^*(p)$, such that the first order conditions

$$\frac{\partial L\big(x^*(p), \lambda^*(p), p\big)}{\partial x} = f'_x\big(x^*(p), p\big) - \sum_{i=1}^{m} \lambda_i^*(p) {g_i}'_x\big(x^*(p), p\big) = 0$$

are satisfied. Assume that $x^* : P \to X$ *and* $\lambda^* : P \to \mathbb{R}^m$ *are* C^1. *Set*

$$V(p) = f\left(x^*(p), p\right).$$

Then V *is* C^1 *and*

$$\frac{\partial V(p)}{\partial p_j} = \frac{\partial L\big(x^*(p), \lambda^*(p), p\big)}{\partial p_j} = \frac{\partial f(x^*, p)}{\partial p_j} - \sum_{i=1}^{m} \lambda_i^*(p) \frac{\partial g_i(x^*, p)}{\partial p_j}.$$

The proof is the same as that of Theorem 6.6.

6.6 The Envelope Theorem and Cost Minimization

minimize $\sum_i w_i x_i$ subject to $f(x_1, \ldots, x_n) = y$

$$L(x, \lambda; y, w) = \sum_i w_i x_i - \lambda\left(f(x_1, \ldots, x_n) - y\right)$$

$$c(y, w) = \sum_i w_i \hat{x}_i(y, w)$$

By the Envelope Theorem,

$$\frac{\partial c}{\partial y} = \left.\frac{\partial L}{\partial y}\right|_{\substack{x=\hat{x}(y,w)\\ \lambda=\hat{\lambda}(y,w)}} = \hat{\lambda}$$

The Lagrange multiplier is the marginal cost. Also,

$$\frac{\partial c}{\partial w_i} = \left.\frac{\partial L}{\partial w_i}\right|_{\substack{x=\hat{x}(y,w)\\ \lambda=\hat{\lambda}(y,w)}} = \hat{x}_i$$

Lecture 7

More about Cost Functions

7.1 Summary of properties of cost functions

Let f be a monotonic production function. The associated cost function $c(w, y)$ is

- continuous
- concave in w
- monotone nondecreasing in (w, y)
- homogeneous of degree one in w, that is, $c(\lambda w, y) = \lambda c(w, y)$ for $\lambda > 0$.

Moreover, if $\hat{x}(w, y)$ is the **conditional factor demand**, then

$$\frac{\partial c(w, y)}{\partial w_i} = \hat{x}_i(w, y).$$

7.2 Cost minimization

Mathematically the cost minimization problem can be formulated as follows.

$$\min_x \; w \cdot x \quad \text{subject to } f(x) \geq y,\ x \geqq 0,$$

where $w \gg 0$ and $y > 0$.

It is clear that if f is monotonic, we may replace the condition $f(x) \geq y$ by $f(x) - y = 0$ without changing the solution. Let $\hat{x}(w, y)$ solve this problem, and assume that $\hat{x} \gg 0$. The

Lagrangian for this minimization problem is

$$w \cdot x - \lambda\big(f(x) - y\big).$$

The gradient of the constraint function (with respect to x) is just $f'(\hat{x})$, which is not zero. Therefore by the Lagrange Multiplier Theorem, there is a Lagrange multiplier $\hat{\lambda}$ (depending on w, y) so that locally the first order conditions

$$w_i - \hat{\lambda}(w,y) f_i\big(\hat{x}(w,y)\big) = 0, \quad i = 1, \dots, n, \tag{7.1}$$

where $f_i(x) = \frac{\partial f(x)}{\partial x_i}$, and the constraint

$$y - f\big(\hat{x}(w,y)\big) = 0 \tag{7.2}$$

hold for all w, y. Note that (7.1) implies that $\hat{\lambda} > 0$.

The second order condition is that

$$\hat{\lambda} \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij}(\hat{x}) v_i v_j \leq 0, \tag{7.3}$$

for all $v \in \mathbb{R}^n$ satisfying

$$f'(\hat{x}) \cdot v = \sum_{i=1}^{n} f_i(\hat{x}) v_i = 0.$$

Using the **method of implicit differentiation** with respect to each w_j on (7.1) yields:

$$\delta_{ij} - \frac{\partial \hat{\lambda}}{\partial w_j} f_i(\hat{x}) - \hat{\lambda} \sum_{k=1}^{n} f_{ik}(\hat{x}) \frac{\partial \hat{x}_k}{\partial w_j} = 0, \quad j \overset{i=1,\dots,n}{=} 1, \dots, n, \tag{7.4}$$

where δ_{ij} is the Kronecker delta,

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Differentiating (7.1) with respect to y yields

$$-\frac{\partial \hat{\lambda}}{\partial y} f_i(\hat{x}) - \hat{\lambda} \sum_{k=1}^{n} f_{ik}(\hat{x}) \frac{\partial \hat{x}_k}{\partial y} = 0, \quad i = 1, \dots, n, \tag{7.5}$$

Now differentiate (7.2) with respect to each w_j to get

$$-\sum_{k=1}^{n} f_k(\hat{x})\frac{\partial \hat{x}_k}{\partial w_j} = 0, \quad j = 1, \dots, n, \tag{7.6}$$

and with respect to y to get

$$-\sum_{k=1}^{n} f_k(\hat{x})\frac{\partial \hat{x}_k}{\partial y} + 1 = 0. \tag{7.7}$$

We can rearrange equations (7.4) through (7.7) into one gigantic matrix equation:

$$\begin{bmatrix} \hat{\lambda} f_{11} & \dots & \hat{\lambda} f_{1n} & f_1 \\ \vdots & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ \hat{\lambda} f_{n1} & \dots & \hat{\lambda} f_{nn} & f_n \\ f_1 & \dots & f_n & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial \hat{x}_1}{\partial w_1} & \dots & \frac{\partial \hat{x}_1}{\partial w_n} & \frac{\partial \hat{x}_1}{\partial y} \\ \vdots & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ \frac{\partial \hat{x}_n}{\partial w_1} & \dots & \frac{\partial \hat{x}_n}{\partial w_n} & \frac{\partial \hat{x}_n}{\partial y} \\ \frac{\partial \hat{\lambda}}{\partial w_1} & \dots & \frac{\partial \hat{\lambda}}{\partial w_n} & \frac{\partial \hat{\lambda}}{\partial y} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 & 0 \\ 0 & \ddots & & & \vdots & \vdots \\ \vdots & & & \ddots & 0 & \vdots \\ 0 & \dots & \dots & 0 & 1 & 0 \\ 0 & \dots & & \dots & 0 & 1 \end{bmatrix}. \tag{7.8}$$

To see where this comes from, break up the $(n+1) \times (n+1)$ matrix equation into four blocks. The upper left $n \times n$ block comes from (7.4). The upper right $n \times 1$ block comes from (7.5). The lower left $1 \times n$ block comes from (7.6), and finally the lower right 1×1 block is just (7.7). This tells us is that

$$\left[\begin{array}{ccc|c} \ddots & & & \vdots \\ & (7.4) & & (7.5) \\ & & \ddots & \vdots \\ \hline \dots & (7.6) & \dots & (7.7) \end{array}\right]$$

Figure 7.1: The blocks in the matrix version of equations (7.4) through (7.7).

$$\begin{bmatrix} \dfrac{\partial \hat{x}_1}{\partial w_1} & \dots & \dfrac{\partial \hat{x}_1}{\partial w_n} & \dfrac{\partial \hat{x}_1}{\partial y} \\ \vdots & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ \dfrac{\partial \hat{x}_n}{\partial w_1} & \dots & \dfrac{\partial \hat{x}_n}{\partial w_n} & \dfrac{\partial \hat{x}_n}{\partial y} \\ \dfrac{\partial \hat{\lambda}}{\partial w_1} & \dots & \dfrac{\partial \hat{\lambda}}{\partial w_n} & \dfrac{\partial \hat{\lambda}}{\partial y} \end{bmatrix} = \begin{bmatrix} \hat{\lambda} f_{11} & \dots & \hat{\lambda} f_{1n} & f_1 \\ \vdots & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ \hat{\lambda} f_{n1} & \dots & \hat{\lambda} f_{nn} & f_n \\ f_1 & \dots & f_n & 0 \end{bmatrix}^{-1}. \tag{7.9}$$

So the second order conditions imply that the $n \times n$ matrix

$$\begin{bmatrix} \frac{\partial \hat{x}_1}{\partial w_1} & \cdots & \frac{\partial \hat{x}_1}{\partial w_n} \\ \vdots & & \vdots \\ \frac{\partial \hat{x}_n}{\partial w_1} & \cdots & \frac{\partial \hat{x}_n}{\partial w_n} \end{bmatrix}$$

is negative semidefinite of rank $n-1$, being the upper left block of the inverse of a bordered matrix that is negative definite under constraint. It follows therefore that

$$\frac{\partial \hat{x}_i}{\partial w_i} \leq 0 \quad i = 1, \ldots, n.$$

Note that this approach provides us with conditions under which the cost function is twice continuously differentiable. It follows from (7.9) that if the bordered Hessian is invertible, the Implicit Function Theorem tells us that $\hat{x}$ and $\hat{\lambda}$ are C^1 functions of w and y (since f is C^2). On the other hand, if $\hat{x}$ and $\hat{\lambda}$ are C^1 functions of w and y, then (7.8) implies that the bordered Hessian is invertible. In either case, the marginal cost $\frac{\partial c}{\partial y} = \hat{\lambda}$, is a C^1 function of w and y, so the cost function is C^2, which is hard to establish by other means.

Returning now to (7.9), note that since the Hessian is a symmetric matrix, we have a number of **reciprocity** results. Namely:

$$\frac{\partial \hat{x}_i}{\partial w_j} = \frac{\partial \hat{x}_j}{\partial w_i} \qquad i = 1, \ldots, n \text{ and } j = 1, \ldots, n$$

and

$$\frac{\partial \hat{x}_i}{\partial y} = \frac{\partial \hat{\lambda}}{\partial w_i} = \frac{\partial^2 c}{\partial w_i \partial y}.$$

7.3 The marginal cost function

Define the cost function c by

$$c(w, y) = \sum_{k=1}^{n} w_k \hat{x}_k(w, y).$$

Then

$$\frac{\partial c(w, y)}{\partial y} = \sum_{k=1}^{n} w_k \frac{\partial \hat{x}_k(w, y)}{\partial y},$$

and

$$\frac{\partial^2 c(w,y)}{\partial y^2} = \sum_{k=1}^{n} w_k \frac{\partial^2 \hat{x}_k(w,y)}{\partial y^2}. \tag{7.10}$$

From (7.1), we have $w_k = \hat{\lambda} f_k(\hat{x})$, so

$$\frac{\partial c(w,y)}{\partial y} = \hat{\lambda} \sum_{k=1}^{n} f_k(\hat{x}) \frac{\partial \hat{x}_k(w,y)}{\partial y} = \hat{\lambda}, \tag{7.11}$$

where the second equality is just (7.7).

That is, *the Lagrange multiplier* $\hat{\lambda}$ *is the marginal cost.*

Now let's see whether the marginal cost is increasing or decreasing as a function of y. Differentiating (7.7) with respect to y yields

$$\sum_{j=1}^{n} \left(\frac{\partial \hat{x}_j}{\partial y} \sum_{i=1}^{n} f_{ij}(\hat{x}) \frac{\partial \hat{x}_i}{\partial y} + f_j(\hat{x}) \frac{\partial^2 \hat{x}_j}{\partial y^2} \right) = 0,$$

or rearranging,

$$\sum_{j=1}^{n} f_j(\hat{x}) \frac{\partial^2 \hat{x}_j}{\partial y^2} = -\sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij}(\hat{x}) \frac{\partial \hat{x}_i}{\partial y} \frac{\partial \hat{x}_j}{\partial y}. \tag{7.12}$$

From (7.10) and (7.1) we have that the left-hand side of (7.12) is $\frac{1}{\hat{\lambda}} \frac{\partial^2 c}{\partial y^2}$. What is the right-hand side?

Fix w and consider the curve $y \mapsto \hat{x}(y)$. This is called an **expansion path**. It traces out the optimal input combination as a function of the level of output. The tangent line to this curve at $\hat{x}$ is just $\{\hat{x} + \alpha v : \alpha \in \mathbb{R}\}$, where

$$v_i = \frac{\partial \hat{x}_i}{\partial y}.$$

Write the output along this tangent line, $f(\hat{x} + \alpha v)$, as a function $\hat{f}$ of α. That is, $\hat{f}(\alpha) = f(\hat{x} + \alpha v)$. By the chain rule,

$$\hat{f}'(\alpha) = \sum_{j=1}^{n} f_j(\hat{x} + \alpha v) v_j,$$

and

$$\hat{f}''(\alpha) = \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij}(\hat{x} + \alpha v) v_i v_j,$$

so

$$\hat{f}''(0) = \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij}(\hat{x}) \frac{\partial \hat{x}_i}{\partial y} \frac{\partial \hat{x}_j}{\partial y}.$$

Thus (7.12) can be written as

$$\frac{\partial^2 c}{\partial y^2} = -\hat{\lambda} \hat{f}''(0).$$

In other words (7.12) asserts that *the slope of the marginal cost curve is increasing (that is, the cost function is a locally convex function of y) when the production function is locally concave on the line tangent to the expansion path*, and vice-versa.

7.4 Average cost and elasticity of scale

Recall that a production function f exhibits **constant returns to scale** if $f(\alpha x) = \alpha f(x)$ for all $\alpha > 0$. It exhibits **increasing returns to scale** if $f(\alpha x) > \alpha f(x)$ for $\alpha > 1$, and **decreasing returns to scale** if $f(\alpha x) < \alpha f(x)$ for $\alpha > 1$. If f is **homogeneous of degree** k, that is, if

$$f(\alpha x) = \alpha^k f(x),$$

then the returns to scale are decreasing, constant, or increasing, as $k < 1$, $k = 1$, or $k > 1$. Define

$$h(\alpha, x) = f(\alpha x).$$

The **elasticity of scale** $e(x)$ of the production function at x is defined to be

$$D_1 h(1, x) \frac{1}{f(x)} = f'(x) \cdot x / f(x),$$

where D_1 denotes the partial derivative with respect to the first argument α:

$$\left. \frac{df(\alpha x)}{d\alpha} \frac{\alpha}{f(x)} \right|_{\alpha=1}.$$

If f is homogeneous of degree k, then $e(x) = k$, as

$$D_1 h(\alpha, x) = k\alpha^{k-1} f(x).$$

Even if f is not homogeneous, following Varian, we can express the elasticity of scale in terms of the marginal and average cost functions, at least for points x that minimize cost uniquely for some (y, w):

$$\begin{aligned}
e\big(\hat{x}(y,w)\big) &= f'(\hat{x}) \cdot \hat{x}/f(\hat{x}) \\
&= f'(\hat{x}) \cdot \hat{x}/y && \text{as } y = f\big(\hat{x}(y,w)\big) \\
&= \frac{w}{\hat{\lambda}} \cdot \hat{x}/y && \text{by the first order condition } w = \hat{\lambda} f'(\hat{x}) \\
&= \frac{c(y,w)/y}{D_y c(y,w)} && \text{as } c(y,w) = w \cdot \hat{x}(y,w), \text{ and by (7.11) } \hat{\lambda} = D_y c(y,w) \\
&= \mathrm{AC}(y)/\mathrm{MC}(y).
\end{aligned}$$

Holding w fixed, and writing the cost simply as a function of y,

$$\frac{d}{dy}\mathrm{AC}(y) = \frac{d}{dy}\frac{c(y)}{y} = \frac{c'(y)y - c(y)}{y^2} = \frac{1}{y}\left(c'(y) - \frac{c(y)}{y}\right) = \frac{1}{y}\left(\mathrm{MC}(y) - \mathrm{AC}(y)\right).$$

Thus

$$\mathrm{AC}'(y) > 0 \Leftrightarrow \mathrm{MC}(y) > \mathrm{AC}(y) \Leftrightarrow e(\hat{x}) < 1.$$

7.5 Average cost and constant returns to scale

If f exhibits constant returns to scale, then:

- the conditional input demand functions $\hat{x}(w, y)$ are homogenous of degree 1 in y.
- Marginal cost = average cost.
- For a price-taking profit maximizer, price = marginal cost = average cost, so profit is zero.
- If price is less than marginal cost, then the optimal output is zero. If price is equal to marginal cost, then every level of output maximizes profit, which is zero. If price is greater than marginal cost, then the profit function is unbounded, so no profit maximizer exists.

7.6 Recovering the Production Function from a Cost Function

We already know from the support function that the input requirement set for y is

$$\{x : w \cdot x \geq c(w, y) \text{ for all } w \in \mathbb{R}^n_+\}.$$

But there is often another way to get a nicer expression for the production function using the envelope theorem.

Example 7.1. Consider the cost function (with two inputs)

$$c(w, y) = y(w_1^\sigma + w_2^\sigma)^{1/\sigma}.$$

By the envelope theorem

$$\frac{\partial c}{\partial w_i} = y\frac{1}{\sigma}(w_1^\sigma + w_2^\sigma)^{\frac{1-\sigma}{\sigma}}\sigma w_i^{\sigma-1} = x_i^*,$$

where $x^*(w, y)$ is the cost minimizing input vector. We can eliminate w_1 and w_2 and solve for y as a function of x_1 and x_2. Here's the trick: exponentiate the above equality to the

$$\rho = \frac{\sigma}{\sigma - 1}$$

power to get

$$y^\rho(w_1^\sigma + w_2^\sigma)^{-1}w_i^\sigma = x_i^\rho$$

and sum over i to get

$$x_1^\rho + x_2^\rho = y^\rho(w_1^\sigma + w_2^\sigma)^{-1}(w_1^\sigma + w_2^\sigma) = y^\rho,$$

which gives the production function

$$y = (x_1^\rho + x_2^\rho)^{1/\rho}.$$

This called the constant elasticity of substitution production function, or the **Arrow–Chenery–Minhas–Solow production function**.

Example 7.2. Given the cost function

$$c(w, y) = y \sum_{i=1}^{n} \alpha_i w_i$$

By the envelope theorem

$$\frac{\partial c}{\partial w_i} = \alpha_i y = x_i^*,$$

where $x^*(w, y)$ is the cost minimizing input vector. This implies that the cost minimizing point x^* is independent of w! Thus

$$y = \frac{x_i^*}{\alpha_i}, \quad i = 1, \ldots, n.$$

Using the support function approach to finding the input requirement set, we see that it is $\{x : x \geqq x^*\}$, so that the production function is

$$y = \min_i \frac{x_i}{\alpha_i}.$$

This sort of production function is a **Leontief production function**.

Lecture 8

Production Functions, Cost Functions, and Production Possibilities

8.1 A Simple Model of Production Possibilities

This is a very simple model of the production possibilities of an economy.

There are n outputs $y_1, \ldots, y_n$ and ℓ factors $v_1, \ldots, v_\ell$. Each output is produced according to the production function $y_j = f^j(v_1^j, \ldots, v_\ell^j)$. There is no joint production, there are no intermediate goods, and there is only one production function for each output.

The supply of factors in the economy are fixed at levels $\omega_1, \ldots, \omega_\ell$.

Assume that for each j, the production function satisfies

$$f^j : \mathbb{R}^l_+ \to \mathbb{R} \text{ is continuous, } C^2 \text{ on } \mathbb{R}^l_{++},\ \nabla f^j \gg 0 \text{ on } \mathbb{R}^l_{++},$$

and that the Hessian

$$D^2 f^j \text{ is negative definite on the subspace orthogonal to } \nabla f^j.$$

You will presently see why we make these assumptions. They guarantee that all the second order conditions hold as strict inequalities.

Production possibility frontier

The **production possibility set (PPS)** is

$$\Big\{y \in \mathbb{R}^n : 0 \leq y^j \leq f^j(v^j),\ v^j \geqq 0,\ j = 1, \dots, n,\ \text{and} \sum_{j=1}^{n} v^j \leqq \omega\Big\},$$

where $\omega = (\omega_1, \dots, \omega_\ell)$ is the factor supply vector. Note that the PPS is compact, since the f^js are continuous and monotonic, so the PPS is the continuous image of the compact set

$$\Big\{(v^1, \dots, v^n) \in \mathbb{R}^{\ell n} : v^j \geqq 0,\ j = 1, \dots, n,\ \text{and} \sum_{j=1}^{n} v^j \leqq \omega\Big\}.$$

The **production possibility frontier** (PPF) is the outer boundary of the PPS.

The PPF solves a constrained maximization problem

The production possibility frontier can be characterized by the following maximization problem.

$$\max_{v^1,\dots,v^n} f^n(v^n) \quad \text{subject to}$$

$$\begin{aligned} f^j(v^j) &= \eta_j, & j &= 1, \dots, n-1 \\ \sum_{j=1}^{n} v_k^j, &= \omega_k & k &= 1, \dots, \ell \\ v_k^j &\geqq 0, & j &= 1, \dots, n \\ & & k &= 1, \dots, \ell. \end{aligned}$$

The Lagrangian is:

$$L(v, \lambda, \mu; \eta, \omega) =$$

$$f^n(v_1^n, \dots, v_\ell^n) + \sum_{j=1}^{n-1} \lambda_j \big(f^j(v_1^j, \dots, v_\ell^j) - \eta_j\big) + \sum_{k=1}^{\ell} \mu_k \Big(\omega_k - \sum_{j=1}^{n} v_k^j\Big).$$

In order to apply the LMT we need to verify that the Lagrange Constraint Qualification is satisfied. That is, we need to show that the gradients of the constraints are linearly independent (at the optimum).

Suppose $\lambda_1, \ldots, \lambda_{n-1}, \mu_1, \ldots, \mu_\ell$ are coefficients on the gradients that yield a linear combination of the gradients that adds up to the zero vector. Then $\mu_1 = \ldots = \mu_\ell = 0$. Thus since each $f^j_k > 0$, we get $\lambda_j = 0$, for all $j = 1, \ldots, n-1$. That is, the gradients are linearly independent.

Thus by the Lagrange multiplier theorem, there are Lagrange multipliers $\hat{\lambda}_j$, $\hat{\mu}_k$, such that the first order conditions are (assuming each $\hat{v}^j_k > 0$) :

$$\hat{\lambda}_j f^j_k(\hat{v}^j) - \hat{\mu}_k = 0 \qquad k \overset{j=1,\ldots,n}{=} 1, \ldots, \ell$$

where for symmetry we define $\hat{\lambda}_n = 1$. This implies

$$\hat{\lambda}_j = \frac{f^n_k}{f^j_k}$$

for any input $k = 1, \ldots, \ell$.

Let $\hat{y}_n(\eta, \omega)$ be the optimal value function. Its graph *is* the PPF. By the envelope theorem, the slope of the PPF satisfies

$$\frac{\partial \hat{y}_n}{\partial \eta_j} = \frac{\partial L}{\partial \eta_j} = -\hat{\lambda}_j = -\frac{f^n_k}{f^j_k}$$

for any $j = 1, \ldots, n-1$, $k = 1, \ldots, \ell$. In other words, λ_j is the marginal opportunity cost of a unit of y_j in terms of y_n.

Also note that

$$\frac{f^j_k}{f^j_{k'}} = \frac{\hat{\mu}_k}{\hat{\mu}_{k'}},$$

which is independent of j. That is, in every industry the slopes of the isoquants are the same.

Second order conditions

While we're at it, let's check the second order conditions. The Hessian of the Lagrangian is the block diagonal $\ell n \times \ell n$ matrix

Let $x = (x^1, \ldots, x^n)$ belong to $\mathbb{R}^{ln}$. The second order condition is that the quadratic form $x'Hx$ is negative semidefinite on the subspace orthogonal to the gradients of the constraints.

It is straightforward to see that this means

$$\sum_{j=1}^{n}\left[\sum_{i=1}^{\ell}\sum_{k=1}^{\ell}\lambda_j f^j_{ik}x^j_i x^j_k\right] \leq 0$$

for all nonzero x satisfying

$$\nabla f^j \cdot x^j = \sum_{k=1}^{\ell} f^j_k x^j_k = 0, \qquad j = 1,\dots,n-1,$$

and

$$\sum_{j=1}^{n} x^j_k = 0 \qquad k = 1,\dots,\ell.$$

What about the case $j = n$? If we can show that $\nabla f^n \cdot x^n = 0$, then by our assumption on the gradients of the f^js, each $\lambda_j > 0$, so by the assumption on the Hessian of the f^js, each bracketed term is nonpositive, and at least one is strictly negative (since at least one $x^j \neq 0$).

To see that $\nabla f^n \cdot x^n = 0$, observe that for each k, $x^j_k = -\sum_{j=1}^{n} x^j_k$. Thus

$$\begin{aligned}
\nabla f^n \cdot x^n &= \sum_{k=1}^{\ell} f^n_k x^n_k \\
&= -\sum_{k=1}^{\ell} f^n_k \sum_{j=1}^{n-1} x^j_k \\
&= -\sum_{j=1}^{n-1}\left[\sum_{k=1}^{\ell}\lambda_j f^j_k x^j_k\right] \\
&= 0.
\end{aligned}$$

The penultimate equality follows from the first order condition that $\lambda_j f^j_k = \mu_k = f^n_k$ for all i.

Relation to cost minimization

Assume that each producer faces the same wages $w = (w_1,\dots,w_\ell)$ for the factors and minimizes costs. To ease notation in this section, I shall suppress the superscripts denoting the particular output.

The cost minimization problem is to

$$\min\ w \cdot v \quad \text{subject to } f(v) \geq y.$$

Form the Lagrangean

$$L(v, \gamma; w, y) = w \cdot v - \gamma\big(f(v) - y\big).$$

The value function is the cost function $c(w, y)$. By the envelope theorem, the marginal cost is

$$\text{MC} = \frac{\partial c}{\partial y} = \frac{\partial L}{\partial y} = \gamma.$$

We also have the first order conditions (check the gradient of the constraint) :

$$w_k - \gamma f_k = 0, \qquad k = 1, \ldots, \ell$$

assuming each $v_k > 0$. (Note that these implies $\gamma > 0$.) In other words,

$$f_k = \frac{w_k}{\text{MC}}$$

Now back to the PPF. If all firms face the same wages and minimize costs, then

$$\frac{\partial \hat{y}_n}{\partial \eta_j} = -\hat{\lambda}_j = -\frac{f_k^n}{f_k^j} = -\frac{\frac{w_k}{\text{MC}_n}}{\frac{w_k}{\text{MC}_j}} = -\frac{MC_j}{MC_n}.$$

That is, the marginal opportunity cost of one unit of y_j expressed in terms of y_n is exactly the ratio of the marginal cost of a unit of y_j (calculated in terms of wages) relative to the marginal cost of a unit of y_n. What this tells us is that marginal costs (derived from wages) indicate real opportunity costs.

8.2 The Averch–Johnson Effect

Averch and Johnson pointed out that a firm subject to rate of return regulation has an incentive not to minimize costs. Thus the apparent cost function for these firms does not yield the true production function. Rate of return regulation is baesed on a couple of Supreme Court rulings:

- *Munn v. Illinois*
- *Hope* case

Maximize $\pi(K)$ subject to $\pi \leq rK$. If the constraint binds the picture looks like figure 8.1. The regulated firm overuses capital in order to get a higher rate base.

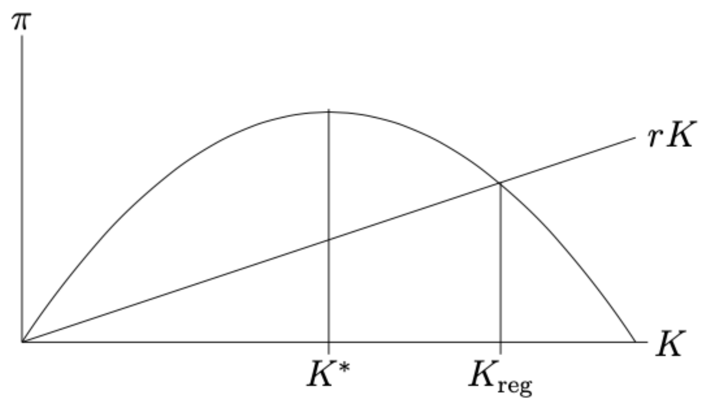


Figure 8.1

8.3 Quasiconcave functions

There are weaker notions of convexity that are commonly applied in economic theory.

Definition 8.1. A function $f : C \to \mathbb{R}$ on a convex subset C of a vector space is:

- **quasiconcave** if for all x, y in C with $x \neq y$ and all $0 < \lambda < 1$

$$f\big(\lambda x + (1-\lambda)y\big) \geq \min\{f(x), f(y)\}.$$

- **strictly quasiconcave** if for all x, y in C with $x \neq y$ and all $0 < \lambda < 1$

$$f\big(\lambda x + (1-\lambda)y\big) > \min\{f(x), f(y)\}.$$

- **explicitly quasiconcave** or **semistrictly quasiconcave** if it is quasiconcave and in addition, for all x, y in C with $x \neq y$ and all $0 < \lambda < 1$

$$f(x) > f(y) \Rightarrow f\big(\lambda x + (1-\lambda)y\big) > \min\{f(x), f(y)\} = f(y).$$

- **quasiconvex** if for all x, y in C with $x \neq y$ and all $0 < \lambda < 1$

$$f\big(\lambda x + (1-\lambda)y\big) \leq \max\{f(x), f(y)\}.$$

- **strictly quasiconvex** if for all x, y in C with $x \neq y$ and all $0 < \lambda < 1$

$$f\big(\lambda x + (1-\lambda)y\big) < \max\{f(x), f(y)\}.$$

- **explicitly quasiconvex** or **semistrictly quasiconvex** if it is quasiconvex and in addition, for all x, y in C with $x \neq y$ and all $0 < \lambda < 1$

$$f(x) < f(y) \Rightarrow f\big(\lambda x + (1-\lambda)y\big) < \max\{f(x), f(y)\} = f(y).$$

There are other choices we could have made for the definition based on the next lemma.

Lemma 8.2. *For a function $f : C \to \mathbb{R}$ on a convex set, the following are equivalent:*

1. *The function f is quasiconcave.*
2. *For each $\alpha \in \mathbb{R}$, the strict upper contour set $[f(x) > \alpha]$ is convex, but possibly empty.*
3. *For each $\alpha \in \mathbb{R}$, the upper contour set $[f(x) \geq \alpha]$ is convex, but possibly empty.*

Proof. (1) $\Rightarrow$ (2) If f is quasiconcave and x, y in C satisfy $f(x) > \alpha$ and $f(y) > \alpha$, then for each $0 \leq \lambda \leq 1$ we have

$$f\big(\lambda x + (1-\lambda)y\big) \geq \min\{f(x), f(y)\} > \alpha.$$

(2) $\Rightarrow$ (3) Note that

$$[f \geq \alpha] = \bigcap_{n=1}^{\infty} [f > \alpha - \tfrac{1}{n}],$$

and recall that the intersection of convex sets is convex.

(3) $\Rightarrow$ (1) If $[f \geq \alpha]$ is convex for each $\alpha \in \mathbb{R}$, then for $y, z \in C$ put $\alpha = \min\{f(y), f(z)\}$ and note that $f\big(\lambda y + (1-\lambda)z\big)$ belongs to $[f \geq \alpha]$ for each $0 \leq \lambda \leq 1$. $\square$

Corollary 8.3. A concave function is quasiconcave. A convex function is quasiconvex.

Lemma 8.4. *A strictly quasiconcave function is also explicitly quasiconcave. Likewise a strictly quasiconvex function is also explicitly quasiconvex.*

Of course, not every quasiconcave function is concave.

Example 8.5 (Explicit quasiconcavity)**.** This example sheds some light on the definition of explicit quasiconcavity. Define $f : \mathbb{R} \to [0,1]$ by

$$f(x) = \begin{cases} 0 & x = 0 \\ 1 & x \neq 0. \end{cases}$$

If $f(x) > f(y)$, then $f\big(\lambda x + (1-\lambda)y\big) > f(y)$ for every $\lambda \in (0,1)$ (since $f(x) > f(y)$ implies $y = 0$). But f is not quasiconcave, as $\{x : f(x) \geq 1\}$ is not convex.

For a proof of the next fact see my notes for Ec 181.

Fact 8.6. Let C be a convex set in $\mathbb{R}^m$. Let f be a lower semicontinuous quasiconcave function on C that has no local maxima. Then f is explicitly quasiconcave.

Theorem 8.7 (Local maxima of explicitly quasiconcave functions)**.** *Let $f : C \to \mathbb{R}$ be an explicitly quasiconcave function (C convex). If x^* is a local maximizer of f, then it is a global maximizer of f over C.*

Proof. Let x belong to C and suppose $f(x) > f(x^*)$. Then by the definition of explicit quasiconcavity, for any $1 > \lambda > 0$, $f\big(\lambda x + (1-\lambda)x^*\big) > f(x^*)$. Since $\lambda x + (1-\lambda)x^* \to x^*$ as $\lambda \to 0$ this contradicts the fact that f has a local maximum at x^*. $\square$

8.4 Quasiconcavity and Differentiability

Quasiconcavity has implications for derivatives.

Proposition 8.8. *Let $C \subset \mathbb{R}^n$ be convex and let $f : C \to \mathbb{R}$ be quasi-concave. Let y belong to C and assume that f has a one-sided directional derivative*

$$f'(x; y-x) = \lim_{\lambda \downarrow 0} \frac{f\big(x + \lambda(y-x)\big) - f(x)}{\lambda}.$$

Then

$$f(y) \geq f(x) \quad \Rightarrow \quad f'(x; y-x) \geq 0.$$

In particular, if f is differentiable at x, then $f'(x) \cdot (y-x) \geq 0$ whenever $f(y) \geq f(x)$.

Proof. If $f(y) \geq f(x)$, then $f\big(x + \lambda(y-x)\big) = f\big((1-\lambda)x + \lambda y\big) \geq f(x)$ for $0 < \lambda \leq 1$ by quasiconcavity. Rearranging implies $\frac{f\left(x+\lambda(y-x)\right)-f(x)}{\lambda} \geq 0$ and taking limits gives the desired result. $\square$

Theorem 8.9. *Let $C \subset \mathbb{R}^n$ be open and let $f : C \to \mathbb{R}$ be quasiconcave and twice-differentiable at $x \in C$. Then*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} D_{i,j}f(x)v_iv_j \leq 0 \qquad \textit{for any } v \textit{ satisfying } f'(x)\cdot v = 0.$$

Proof. Pick $v \in \mathbb{R}$ and define

$$g(\lambda) = f(x+\lambda v).$$

Then

$$g(0) = f(x), \qquad g'(0) = f'(x)\cdot v, \qquad g''(0) = \sum_{i=1}^{n}\sum_{j=1}^{n} D_{i,j}f(x)v_iv_j.$$

What we have to show is that if $g'(0) = 0$, then $g''(0) \leq 0$. Assume for the sake of contradiction that $g'(0) = 0$ and $g''(0) > 0$. Then g has a strict local minimum at zero. That is, for $\epsilon > 0$ small enough, $f(x+\epsilon v) > f(x)$ and $f(x-\epsilon v) > f(x)$. But by quasiconcavity,

$$f(x) = f\big(\tfrac{1}{2}(x+\epsilon v) + \tfrac{1}{2}(x-\epsilon v)\big) \geq \min\{f(x+\epsilon v), f(x-\epsilon v)\} > f(x),$$

a contradiction. ∎

Lecture 9

Quasiconcavity and Maximization

9.1 Maximization and Inequality Constraints

Theorem 9.1 (Karush–Kuhn–Tucker). *Let $f, g_1, \ldots, g_m : \mathbb{R}^n_+ \to \mathbb{R}$ be differentiable at x^*, and let x^* be a constrained local maximizer of f subject to $g(x) \geqq 0$ and $x \geqq 0$.*

Let $B = \{i : g_i(x^*) = 0\}$ denote the set of binding functional constraints, and let $Z = \{j : x_j = 0\}$ denote the set of binding nonnegativity constraints on the variables. Assume that x^* satisfies the Kuhn–Tucker Constraint Qualification (see section 9.3 below). Then there exists $\lambda^* \in \mathbb{R}^m$ such that

$$f'(x^*) + \sum_{i=1}^{m} \lambda_i^* {g_i}'(x^*) \leqq 0,$$

$$x^* \cdot \left(f'(x^*) + \sum_{i=1}^{m} \lambda_i^* {g_i}'(x^*) \right) = 0,$$

$$\lambda^* \geqq 0,$$

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

9.2 Karush–Kuhn–Tucker Theorem for Minimization

Theorem 9.2 (Karush–Kuhn–Tucker). *Let $f, g_1, \ldots, g_m : \mathbb{R}^n_+ \to \mathbb{R}$ be differentiable at x^*, and let x^* be a constrained local minimizer of f subject to $g(x) \geqq 0$ and $x \geqq 0$.*

Let $B = \{i : g_i(x^*) = 0\}$, and let $Z = \{j : x_j^* = 0\}$,. Assume that x^* satisfies the

Kuhn–Tucker Constraint Qualification. Then there exists $\lambda^* \in \mathbb{R}^m$ such that

$$f'(x^*) - \sum_{i=1}^{m} \lambda_i^* {g_i}'(x^*) \geqq 0,$$

$$x^* \cdot \left(f'(x^*) - \sum_{i=1}^{m} \lambda_i^* {g_i}'(x^*) \right) = 0,$$

$$\lambda^* \geqq 0,$$

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

Proof. Minimizing f is the same as maximizing $-f$. The Kuhn–Tucker conditions for this imply that there exists $\lambda^* \in \mathbb{R}^m_+$ such that

$$-f'(x^*) + \sum_{i=1}^{m} \lambda_i^* {g_i}'(x^*) \leqq 0,$$

and the conclusion follows by multiplying this by -1. □

9.3 Constraint Qualifications

Definition 9.3. Let $f, g_1, \ldots, g_m : \mathbb{R}^n_+ \to \mathbb{R}$. Let

$$C = \{x \in \mathbb{R}^n : x \geqq 0,\ g_i(x) \geq 0,\ i = 1, \ldots, m\}.$$

denote the **constraint set**. Consider a point $x^* \in C$ and let

$$B = \{i : g_i(x^*) = 0\} \text{ and } Z = \{j : x_j^* = 0\},$$

index the set of binding functional constraints and the set of binding nonnegativity constraints at x^*. The point x^* satisfies the **Kuhn–Tucker Constraint Qualification** if $f, g_1, \ldots, g_m$ are differentiable at x^*, and for every $v \in \mathbb{R}^n$ satisfying

$$\begin{aligned} v_j = v \cdot e^j \geq 0 \qquad & j \in Z, \\ v \cdot {g_i}'(x^*) \geq 0 \qquad & i \in B, \end{aligned}$$

there is a continuous curve $\xi\colon [0,\epsilon) \to \mathbb{R}^n$ satisfying

$$\begin{aligned} \xi(0) &= x^*, \\ \xi(t) \in C & \quad \text{for all } t \in [0,\epsilon), \\ D\xi(0) &= v, \end{aligned}$$

where $D\xi(0)$ is the one-sided directional derivative at 0.

This condition is actually a little weaker than Kuhn and Tucker's condition. They assumed that the functions $f, g_1, \dots, g_m$ were differentiable everywhere and required ξ to be differentiable everywhere. You can see that it may be difficult to verify it in practice.

To better understand the hypotheses of the theorem, let's look at a classic example of its failure.

Example 9.4 (Failure of the Kuhn–Tucker Constraint Qualification)**.** Let $f : \mathbb{R}^2 \to \mathbb{R}$ via $f(x, y) = x$ and $g : \mathbb{R}^2 \to \mathbb{R}$ via $g(x, y) = (1 - x)^3 - y$. The curve $g = 0$ is shown in figure 9.1, and the constraint set in figure 9.2.

Clearly $(x^*, y^*) = (1, 0)$ maximizes f subject to $(x, y) \geqq 0$ and $g \geq 0$. At this point we have $g'(1, 0) = (0, -1)$ and $f' = (1, 0)$. Note that no λ (nonnegative or not) satisfies

$$(1, 0) + \lambda(0, -1) \leqq (0, 0).$$

Fortunately for the theorem, the Constraint Qualification fails at $(1, 0)$. To see this, note that the constraint $g \geq 0$ binds, that is $g(1, 0) = 0$ and the second coordinate of (x^*, y^*) is zero. Suppose $v = (v_x, v_y)$ satisfies

$$v \cdot g'(1, 0) = v \cdot (0, -1) = -v_y \geq 0 \qquad \text{and} \qquad v \cdot e^2 = v_y \geq 0,$$

that is, $v_y = 0$. For instance, take $v = (1, 0)$. The constraint qualification requires that there is a path starting at $(1, 0)$ in the direction $(1, 0)$ that stays in the constraint set. Clearly no such path exists, so the constraint qualification fails.

The next result, which may be found in Arrow, Hurwicz, and Uzawa, provides a tractable sufficient condition for the KTCQ.

Theorem 9.5 (Constraint Qualifications)**.** *In theorem 9.1, the KTCQ may be replaced by any of the conditions below.*

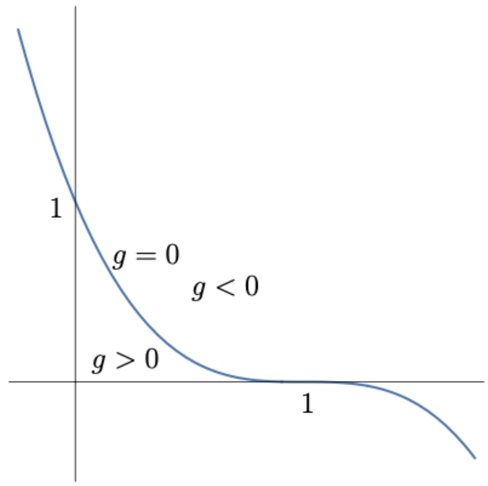


Figure 9.1: The function $g(x, y) = (1 - x)^3 - y$.

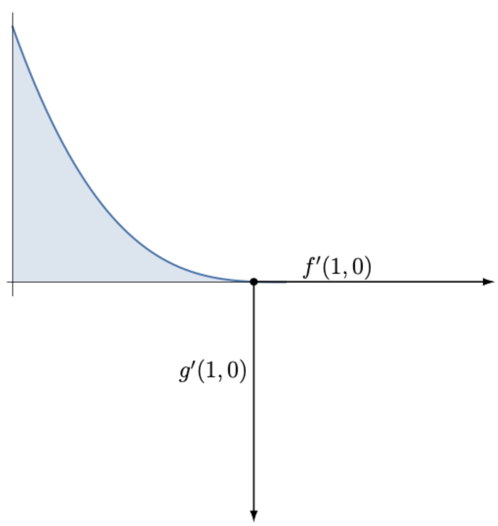


Figure 9.2: This constraint set violates the Constraint Qualification.

1. *Each g_i is convex. (This includes the case where each is linear.)*

2. *Each g_i is concave and there exists some $\hat{x} \gg 0$ for which each $g_i(\hat{x}) > 0$.*

3. *The set $\{e^j : j \in Z\} \cup \{g_i{}'(x^*) : i \in B\}$ is linearly independent.*

9.4 Cost Function for Linear Production Function

With this constant returns to scale production function, all inputs are perfect substitutes for each other (provided units are chosen properly).

$$y = \alpha_1 x_1 + \ldots + \alpha_n x_n$$

where each $\alpha_i > 0$, $i = 1, \ldots, n$.

The Lagrangian for the cost minimization problem is

$$\sum_{i=1}^{n} w_i x_i - \lambda \left(\sum_{i=1}^{n} \alpha_i x_i - y \right)$$

and the naïve first order conditions are

$$\frac{\partial L}{\partial x_i} = w_i - \lambda \alpha_i = 0 \qquad i = 1, \ldots, n,$$

which taken at face value imply $\dfrac{w_1}{\alpha_1} = \cdots = \dfrac{w_n}{\alpha_n}$, which is unlikely since these are all exogenous. This is a red flag that signals that the nonnegativity constraints are binding and that you need to examine the Kuhn–Tucker first order conditions. They are

$$w_i - \lambda \alpha_i \geq 0 \qquad i = 1, \ldots, n,$$

and

$$x_i > 0 \Rightarrow w_i - \lambda \alpha_i = 0 \qquad \text{and} \qquad w_i - \lambda \alpha_i > 0 \Rightarrow x_i = 0.$$

In addition, $\lambda \geq 0$ and $\lambda \left(\sum_{i=1}^{n} \alpha_i x_i - y \right) = 0$.

Thus

$$\frac{w_i}{\alpha_i} \geq \lambda \qquad i = 1, \ldots, n.$$

The question is, can we have strict inequality for each i? The answer is no, as that would imply $x_i = 0$ for each i and the output would be zero, not $y > 0$. So the solution must satisfy

$$\hat{\lambda} = \min_i \frac{w_i}{\alpha_i}.$$

Let i^* satisfy $\hat{\lambda} = \frac{w_{i^*}}{\alpha_{i^*}}$. That is, i^* is a factor that maximizes "bang per buck." Then the conditional factor demand given by:

$$\hat{x}_i = \begin{cases} \dfrac{y}{\alpha_i}, & i = i^* \\ 0, & i \neq i^* \end{cases}$$

minimizes cost, and the cost function is

$$c(y, w) = y \cdot \min\left\{\frac{w_1}{\alpha_1}, \cdots, \frac{w_n}{\alpha_n}\right\}.$$

This is the cost function even if i^* is not unique, but when there is more than one such i^*, the conditional factor demand is no longer a unique input vector, but rather a set of cost minimizing input vectors. In fact, the set of cost minimizing input vectors is the convex set:

$$\operatorname{co}\left\{\frac{y}{\alpha_i}e^i : \frac{w_i}{\alpha_i} = \hat{\lambda} = \min_j \frac{w_j}{\alpha_j}\right\}.$$

Note that even though the production function is very smooth, the cost function fails to be differentiable (for $n \geq 2$). This is to be expected since the bordered Hessian of the production function is given by

$$\begin{bmatrix} f_{11} & \dots & f_{1n} & f_1 \\ \vdots & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ f_{n1} & \dots & f_{nn} & f_n \\ f_1 & \dots & f_n & 0 \end{bmatrix} = \begin{bmatrix} 0 & \dots & 0 & \alpha_1 \\ \vdots & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ 0 & \dots & 0 & \alpha_n \\ \alpha_1 & \dots & \alpha_n & 0 \end{bmatrix},$$

which is singular for $n \geq 2$. (It has rank 2.)

9.5 Saddle point Theorem

Definition 9.6. Let $\phi : X \times Y \to \mathbb{R}$. A point (x^*, y^*) in $X \times Y$ is a **saddle point of ϕ (over $X \times Y$)** if it satisfies

$$\phi(x, y^*) \leq \phi(x^*, y^*) \leq \phi(x^*, y) \quad \text{for all} \quad x \in X,\ y \in Y.$$

Definition 9.7. Given $f, g_1, \ldots, g_m : C \to \mathbb{R}$, the associated **Lagrangian** $L : C \times \Lambda \to \mathbb{R}$ is defined by

$$L(x, \lambda) = f(x) + \sum_{j=1}^{m} \lambda_j g_j(x) = f(x) + \lambda \cdot g(x),$$

where Λ is an appropriate subset of $\mathbb{R}^m$. (Usually $\Lambda = \mathbb{R}^m$ or $\mathbb{R}^m_+$.) The components of λ are called **Lagrange multipliers**.

Theorem 9.8. *Let $C \subset \mathbb{R}^n$ be convex, and let $f, g_1, \ldots, g_m : C \to \mathbb{R}$ be concave. Assume in addition that* ***Slater's Condition****,*

$$\exists \bar{x} \in C \quad g(\bar{x}) \gg 0, \tag{S}$$

is satisfied. Then x^* maximizes f subject to the constraints $g_j(x) \geq 0$, $j = 1, \ldots, m$ if and only if

there exist real numbers $\lambda_1^*, \ldots, \lambda_m^* \geq 0$ such that $x^*, \lambda_1^*, \ldots, \lambda_m^*$ is a saddle point of the Lagrangian for $x \in C$, $\lambda \geqq 0$. That is,

$$L(x, \lambda^*) \leq L(x^*, \lambda^*) \leq L(x^*, \lambda) \quad x \in C, \quad \lambda \geqq 0, \tag{9.1}$$

where $L(x, \lambda) = f(x) + \lambda \cdot g(x)$.

Furthermore, in this case

$$\sum_{j=1}^{m} \lambda_j^* g_j(x^*) = 0. \tag{9.2}$$

In other words, for a concave programming problem, the optimal x^* maximizes the Lagrangian $L(\cdot, \lambda^*)$. The role of the Lagrange multipliers is to provide conversion factors or prices to convert a constrained maximization problem to an unconstrained maximization problem.

The next example shows what can go wrong when Slater's Condition fails.

Example 9.9. In this example, due to Slater, $C = \mathbb{R}$, $f(x) = x$, and $g(x) = -(1-x)^2$. Note that Slater's Condition fails because $g \leq 0$. The constraint set $[g \geq 0]$ contains only 1. Therefore f attains a constrained maximum at 1. There is however no saddle point at all of the Lagrangian

$$L(x, \lambda) = x - \lambda(1-x)^2 = -\lambda + (1+2\lambda)x - \lambda x^2.$$

To see this, observe the first order condition for a maximum in x is $\frac{\partial L}{\partial x} = 0$, or $1+2\lambda-2\lambda x = 0$, which implies $x > 1$ since $\lambda \geq 0$. But for $x > 1$, $\frac{\partial L}{\partial \lambda} = -(1-x)^2 < 0$, so no minimum with respect to λ exists.

9.5.1 The role of Slater's Condition

In this section we present a geometric argument that illuminates the role of Slater's Condition in the saddle point theorem. Let us consider the argument underlying its proof. In the framework of theorem 9.8, define the function $h : C \to \mathbb{R}^{m+1}$ by

$$h(x) = \big(g_1(x), \ldots, g_m(x), f(x) - f(x^*)\big)$$

and set

$$H = \{h(x) : x \in C\} \quad \text{and} \quad \hat{H} = \big\{y \in \mathbb{R}^{m+1} : \exists x \in C \quad y \leqq h(x)\big\}.$$

Then $\hat{H}$ is a convex set bounded in part by H. Figure 9.3 depicts the sets H and $\hat{H}$ for Slater's example 9.9, where $f(x) - f(x^*)$ is plotted on the vertical axis and $g(x)$ is plotted on the horizontal axis. Now if x^* maximizes f over the convex set C subject to the constraints $g_j(x) \geq 0$, $j = 1, \ldots, m$, then $h(x^*)$ has the largest vertical coordinate among all the points in H whose horizontal coordinates are nonnegative.

The semipositive $m+1$-vector $\hat{\lambda}^* = (\lambda_1^*, \ldots, \lambda_m^*, \mu^*)$ from theorem 9.8 is obtained by separating the convex set $\hat{H}$ and $\mathbb{R}^{m+1}_{++}$. It has the property that

$$\hat{\lambda}^* \cdot h(x) \leq \hat{\lambda}^* h(x^*)$$

for all $x \in C$. That is, the vector $\hat{\lambda}^*$ defines a hyperplane through $h(x^*)$ such that the entire set $\hat{H}$ lies in one half-space. It is clear in the case of Slater's example that the hyperplane is a vertical line, since it must be tangent to H at $h(x^*) = (0, 0)$. The fact that the hyperplane is vertical means that μ^* (the multiplier on f) must be zero.

If there is a non-vertical hyperplane through $h(x^*)$, then μ^* is nonzero, so we can divide by it and obtain a full saddle point of the true Lagrangian. This is where Slater's condition

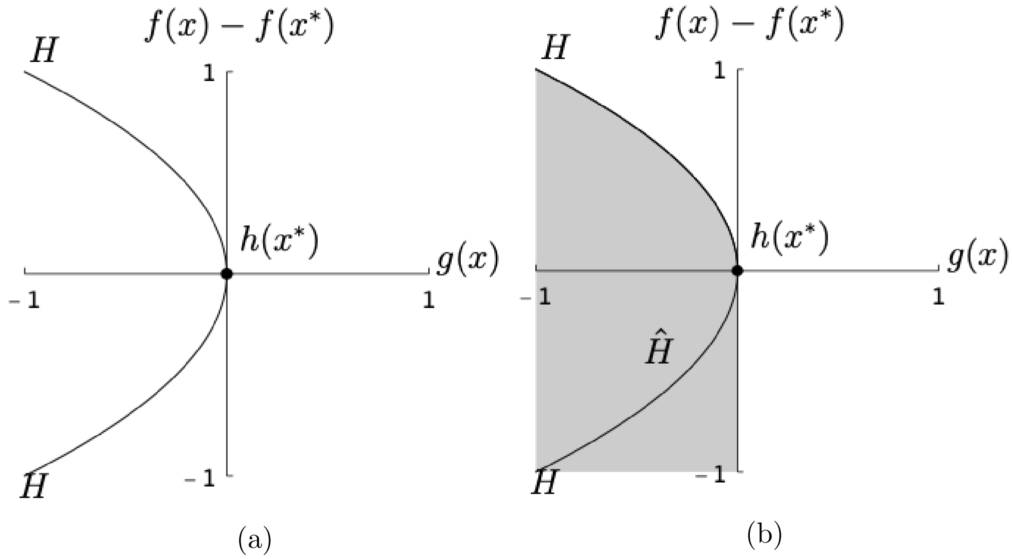


Figure 9.3: The sets H and $\hat{H}$ for Slater's example.

comes in.

In the one dimensional, one constraint case, Slater's Condition reduces to the existence of $\bar{x}$ satisfying $g(\bar{x}) > 0$. This rules out having a vertical supporting line through x^*. To see this, note that the vertical component of $h(x^*)$ is $f(x^*) - f(x^*) = 0$. If $g(x^*) = 0$, then the vertical line through $h(x^*)$ is simply the vertical axis, which cannot be, since $h(\bar{x})$ lies to the right of the axis. If $g(x^*) > 0$, then $\hat{H}$ includes every point below $h(x^*)$, so the only line separating $\hat{H}$ and $\mathbb{R}^2_{++}$ is horizontal, not vertical. See figure 9.4. In Figure 9.4, the shaded

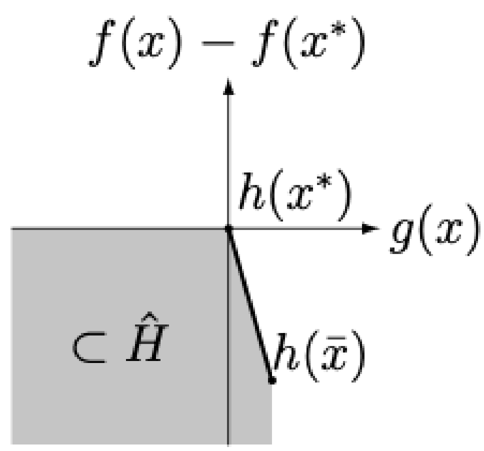


Figure 9.4: Slater's condition guarantees a non-vertical supporting line.

area is included in $\hat{H}$. For instance, let $C = (-\infty, 0]$, $f(x) = x$, and $g(x) = x + 1$. Then the set $\hat{H}$ is just $\{y \in \mathbb{R}^2 : y \leqq (0,1)\}$.

If f and the g_js are linear, then Slater's Condition is not needed to guarantee a non-vertical supporting line. Intuitively, the reason for this is that for the linear case, the set $\hat{H}$ is polyhedral, so even if $g(x^*) = 0$, there is still a non-vertical line separating $\hat{H}$ and $\mathbb{R}^m_{++}$. The proof of this fact relies on results about linear inequalities. It is subtle because Slater's condition rules out a vertical supporting line. In the linear case, there may be a vertical supporting line, but if there is, there is also a non-vertical supporting line that yields a Lagrangian saddle point. As a case in point, consider $C = (-\infty, 0]$, $f(x) = x$, and $g(x) = x$. Then the set $\hat{H}$ is just $\{y \in \mathbb{R}^2 : y \leqq 0\}$, which is separated from $\mathbb{R}^2_{++}$ by every semipositive vector.

9.6 Lagrange Multipliers and Decentralization

Recall that the **production possibility set** (PPS) is

$$\Big\{y \in \mathbb{R}^n : 0 \leq y^j \leq f^j(v^j),\ v^j \geqq 0,\ j = 1, \dots, n,\ \text{and } \sum_{j=1}^n v^j \leqq \omega\Big\}.$$

Note that the PPS is compact, since the f^js are assumed to be continuous and monotonic, so the PPS is the continuous image of the compact set

$$\Big\{(v^1, \dots, v^n) \in \mathbb{R}^{ln} : v^j \geqq 0,\ j = 1, \dots, n,\ \text{and } \sum_{j=1}^n v^j \leqq \omega\Big\}.$$

The **production possibility frontier** (PPF) is the outer boundary of the PPS. That is, y belongs to the PPF if y belongs to the PPS and there is no y' in the PPS distinct from y with $y' \geqq y$. Such a y is also called **technically efficient**. It is easy to verify that if each f^j is concave, then the PPS is convex. Therefore every point on the PPF is a support point. That is, if y belongs to the PPF, then there is a vector p of strictly positive prices such that y maximizes p over the PPS. This follows from the separating hyperplane theorem applied to the PPS and $\{z : z \gg y\}$. In this case the PPF can be parameterized by p.

Maximizing the value of output:

$$\max_{v^1,\dots,v^n} \sum_{j=1}^{n} p_j f^j(v^j) \quad \text{subject to}$$

$$\sum_{j=1}^{n} v_k^j \leq \omega_k \quad k = 1,\dots,\ell$$

$$v_k^j \geq 0 \qquad j = 1,\dots,n$$

$$k = 1,\dots,\ell.$$

The Lagrangian is:

$$L(v,\mu) = \sum_{j=1}^{n} p_j f^j(v^j) + \sum_{k=1}^{\ell} \mu_k \left(\omega_k - \sum_{j=1}^{j} v_k^j \right).$$

Note that as long as each $\omega_k > 0$, then Slater's Condition is satisfied. So a point v^* solves the maximization problem if and only if the there are λ^* and μ^* such that $(v^*;\lambda^*,\mu^*)$ is a saddle point of the Lagrangian.

Let's examine a simplified version with $n = 2$ and $\ell = 2$, and let's further name the inputs labor, L, and capital, K, available in fixed quantities $\bar{L}$ and $\bar{K}$, and let us also use w and r for the Lagrange multipliers instead of μ. (The same argument works in the general case—there is just more notation.) The saddle point condition is

$$\begin{aligned} & p_1 f^1(L_1,K_1) + p_2 f^2(L_2,K_2) + w^*(\bar{L} - L_1 - L_2) + r^*(\bar{K} - K_1 - K_2) \\ \leq\ & p_1 f^1(L_1^*,K_1^*) + p_2 f^2(L_2^*,K_2^*) + w^*(\bar{L} - L_1^* - L_2^*) + r^*(\bar{K} - K_1^* - K_2^*) \end{aligned} \tag{9.3}$$

$$\begin{aligned} & p_1 f^1(L_1^*,K_1^*) + p_2 f^2(L_2^*,K_2^*) + w^*(\bar{L} - L_1^* - L_2^*) + r^*(\bar{K} - K_1^* - K_2^*) \\ \leq\ & p_1 f^1(L_1^*,K_1^*) + p_2 f^2(L_2^*,K_2^*) + w(\bar{L} - L_1^* - L_2^*) + r(\bar{K} - K_1^* - K_2^*) \end{aligned} \tag{9.4}$$

where (9.3) holds for all $(L_1,L_2,K_1,K_2) \geqq 0$ and (9.4) holds for all $(\lambda,w,r) \geqq 0$. Evaluating (9.3) at $L_1 = L_1^*$ and $K_1 = K_1^*$, and canceling common terms yields

$$p_2 f^2(L_2,K_2) - w^* L_2 - r^* K_2 \leq p_2 f^2(L_2^*,K_2^*) - w^* L_2^* - r^* K_2^*$$

This says that (L_2^*,K_2^*) maximizes profit at price p_2 and wages w^* and rental rate r^*.

Similarly, evaluating (9.3) at $L_2 = L_2^*$ and $K_2 = K_2^*$, and canceling common terms yields

$$p_1 f^1(L_1, K_1) - w^* - L_1 - r^* K_1 \leq p_1 f^1(L_1^*, K_1^*) - w^* L_1^* - r^* K_1^*$$

which says that (L_1^*, K_1^*) maximizes profit at price p_1 and wages w^* and rental rate r^*.

In other words,

the Lagrange multipliers on the resource constraints are prices that decentralize the problem of maximizing the value of output.

But the saddle point theorem also tells us we can go backwards! That is, if we maximize profits given wages and the resource markets clear, then profit maximization leads to maximization of output value.

That is, suppose (L_i^*, K_i^*) maximizes $p_i f^i(L, K) - wL - rK$, for $i = 1, 2$, and assume that $K_1^* + K_2^* = \bar{K}$ and $L_1^* + L_2^* = \bar{L}$. Then we have

$$\begin{aligned} &p_1 f^1(L_1, K_1) - wL_1 - rK_1 + p_2 f^2(L_2, K_2) - wL_2 - rK_2 \\ &\qquad \leq p_1 f^1(L_1^*, K_1^*) - wL_1^* - rK_1^* + p_2 f^2(L_2^*, K_2^*) - wL_2^* - rK_2^* \end{aligned}$$

for all L_1, K_1, L_2, K_2. Add $w\bar{L} + r\bar{K}$ to both sides and rearrange to get

$$\begin{aligned} &p_1 f^1(L_1, K_1) + p_2 f^2(L_2, K_2) + w(\bar{L} - L_1 - L_2) + r(\bar{K} - K_1 - K_2) \\ &\qquad \leq p_1 f^1(L_1^*, K_1^*) + p_2 f^2(L_2^*, K_2^*) + w(\bar{L} - L_1^* - L_2^*) + r(\bar{K} - K_1^* - K_2^*) \end{aligned}$$

for all L_1, K_1, L_2, K_2. This is the first saddle point inequality. The second saddle point inequality is

$$\begin{aligned} &p_1 f^1(L_1^*, K_1^*) + p_2 f^2(L_2^*, K_2^*) + w(\bar{L} - L_1^* - L_2^*) + r(\bar{K} - K_1^* - K_2^*) \\ &\qquad \leq p_1 f^1(L_1^*, K_1^*) + p_2 f^2(L_2^*, K_2^*) + w'(\bar{L} - L_1^* - L_2^*) + r'(\bar{K} - K_1^* - K_2^*) \end{aligned}$$

for all $(w', r') \geqq (0, 0)$, which is true since the resource constraints are assumed to bind.

Lecture 10

Introduction to Capital Theory

10.1 Present discounted value

Comparing income streams

If you can invest one dollar at an annual rate of interest r, then in one year you will have $1 + r$ dollars. If that $1 + r$ is reinvested for another year it will be worth $(1 + r)^2$ dollars in two years. In general, after t years of reinvestment, it will be worth $(1 + r)^n$ dollars.

Equivalently, to get one dollar in t years, you need to invest $1/(1 + r)^t$ dollars today and reinvest the proceeds annually. In this sense, $1/(1 + r)^t$ dollars today is worth 1 dollar in t years, and is called the **present discounted value** of \$1 at date t.

Moreover present value is linear: In order to have x_t dollars at each date t, $t = 1, \ldots, n$, you need to invest

$$\sum_{t=1}^{n} \frac{x_t}{(1 + r)^t}$$

today ($t = 0$). This is the present discounted value of the **income stream** $x_1, \ldots, x_n$. If the sequence x_t is bounded (or does not grow too fast), then we can compute the present value of the stream $x_1, x_2, \ldots$ as a convergent infinite series.

The present value of an income stream $\mathbf{x}$ is what you should be willing to pay today to receive that income stream. In fact, if two income streams $\mathbf{x} = (x_1, x_2, \ldots)$ and $\mathbf{y} = (y_1, y_2, \ldots)$ have the same present value, then you can convert $\mathbf{y}$ into $\mathbf{x}$ via a series of borrowing and investing transactions, where you can borrow and invest at the same rate r. (If there are no financial intermediaries, borrowers pay investors, so the rate on investing and borrowing must be equal.) Equivalently, the present value of $\mathbf{x}$ is the stock of cash you have to have today in order to receive the income stream $\mathbf{x}$.

Compounding periods

Interest is often compounded more frequently than annually. If interest is compounded n times annually, the annual interest rate is divided by n and credited at the end of each $1/n$-th year, so the value in t years is $(1 + \frac{r}{n})^{nt}$. Now

$$\lim_{n\to\infty}\left(1+\frac{r}{n}\right)^{nt} = e^{rt},$$

so with **continuous compounding** the **present discounted values** of a dollar at time t is just e^{-rt} today.

Continuous income flows

Sometimes economists prefer to work with continuous time instead of discrete periods. In this case, we need to distinguish between **stocks** and **flows**. The simplest analogy is that of filling a swimming pool with with a garden hose. Water flows from the hose at variable rate and creates a stock of water in the pool. Flows are measured in gallons per minute (or other appropriate units such as liters per second) and over the course of an interval of time the integrated flow becomes a stock, which is measured in gallons (or liters). In the swimming pool case, the stock cannot change instantaneously, they must change as the result of inflows or outflows over an interval of time. But in economics and finance sometimes the stock of capital can change discontinuously by borrowing or lending part of the stock.

We shall measure time so that "now" is $t = 0$. Consider now a "flow" of income $f(t)$ at time t measured in dollars per second (or euros per second), for $t \geq 0$. The present discounted value of the flow f is the stock of cash

$$\int_0^\infty f(t)e^{-rt}\,dt.$$

The stock of cash is in units of dollars (or euros).

See the Appendix for a generalization of this to time-varying interest rates.

10.2 A typical investment problem

The following notes are based on the wonderful book *Income, Wealth, and the Maximum Principle* by Martin L. Weitzman and the paper by Robert Dorfman. A small firm has a **capital stock** (measured in dollars) that it uses to produce a *flow* of **income**. Let $f(K)$

denote the flow rate of gross income that the firm produces using a capital stock of size K, and assume that f is twice continuously differentiable, strictly concave, and strictly increasing with $f' > 0$ and $f'' < 0$.

We assume that capital depreciates at the constant rate $\delta \geq 0$. That is, the outflow rate of capital δK. The firm can increase its capital stock by saving a flow of I of the income as a net **investment** in the capital stock.

The firm can borrow and lend at the market interest rate $r > 0$, so what it cares about is the present discounted value of its net income. What the firm must choose is a time path, or **trajectory** of its **control variable** I (net investment) over the time horizon $[0, \infty)$. (In what follows I shall use bold letters to denote trajectories.) The control influences the **state variable** K through the differential "**equation of motion**"

$$\dot{K}(t) = I(t).$$

(Here I use the traditional dot notation to indicate derivatives with respect to time.) For simplicity assume that there is a maximal rate $\bar{I}$ of investment, and note that with no gross investment the net investment rate is $I(t) = -\delta K(t)$ due to depreciation. Then the firm faces constraints given by the initial condition

$$K(0) = K_0,$$

a nonnegativity constraint $K(t) \geq 0$ for all t, and

$$-\delta K(t) \leq I(t) \leq \bar{I}.$$

Its goal is to maximize

$$\int_0^\infty \left[f\big(K(t)\big) - I(t)\right] e^{-rt}\, dt.$$

You might think that solving this requires a manager who is very far-sighted and can balance the trade-off between investing more now at the expense of current income to provide more income in the future, but in fact

> there is a trajectory $\mathbf{p}$ of "prices" that temporally decentralize this problem so that each instant t the manager chooses the level of investment $I(t)$ to maximize a simple function of K, I, and p, called the **Hamiltonian**. The entire trade-off is summarized at each point in time by the value $\mathbf{p}(t)$.

10.3 A more general mathematical formulation

This is a special case of the following maximization problem,

$$\max_I \int_0^\infty G\big(K(t), I(t)\big) e^{-rt}\, dt$$

subject to the constraints (10.1)–(10.4).

$$\dot{K}(t) = I(t). \tag{10.1}$$

$$K(0) = K_0. \tag{10.2}$$

$$m\big(K(t)\big) \leq I(t) \leq M\big(K(t)\big) \qquad t \geq 0, \tag{10.3}$$

$$K(t) \geq 0 \qquad t \geq 0. \tag{10.4}$$

Here m and M are known functions that of course satisfy $m(K) \leq M(K)$ for all $K \geq 0$.[1] We shall require that m be convex and decreasing and M be concave and increasing. For many problems m is the zero function.

The function G is the instantaneous **gain function**, and r is the **discount rate**. The arguments of G are the current levels of the **state variable** K, and the **control variable** I.

Assumption 10.1. *G is concave and continuously differentiable, and satisfies $G(0, 0) = 0$, and $D_1G \geq 0$ and $D_2G < 0$ everywhere.*

Admissible controls

We restrict attention to control trajectories that are piecewise continuous. In other words, I is required to have at most finitely many discontinuities in any finite time interval.

10.4 Steady states

A **steady state** is a pair (K, I) of trajectories satisfying

$$K(t) = K, \quad I(t) = 0 \;\; \forall t \geq 0.$$

[1] The general formulation of the Maximum Principle allows for additional constraints.

We shall refer to a steady state by the level K of the capital stock it maintains. A steady state may or may not exist, and if it exists, it may or may not be optimal. So fix $K > 0$ and let

$$\phi_0 = \int_0^\infty G(K,0)e^{-rt}\,dt = G(K,0)/r,$$

the present value of the steady state K. A standard technique from the calculus of variations is to look at trajectory and consider a variation on it. The variation on $I = 0$ that I want to consider is this. Invest at the rate ϵ/δ for a short time δ to increase the capital stock to $K + \epsilon$. Intuitively, this incurs an "instantaneous" cost on the order of $D_2G(K,0)\epsilon$ now, but provides an increment in the present discounted value of the flow of $D_1G(K,0)\epsilon/r$. Thus it will pay to adjust the capital stock up or down by ϵ unless $-D_1G(K,0)/D_2G(K,0) = r$. Weitzman defines the **stationary return on capital** by

$$R(K) = -\frac{D_1G(K,0)}{D_2G(K,0)}.$$

So if a steady state $K > 0$ is optimal, then

$$R(K) = -\frac{D_1G(K,0)}{D_2G(K,0)} = r \tag{10.5}$$

must necessarily hold.

Related functions

We now define three functions related to the problem above. The first is the **value function** V. It is the maximized value of the objective function as a function of the initial capital stock. That is,

$$V(K) = \max_I \int_0^\infty G\big(K(t), I(t)\big)e^{-rt}\,dt$$

where the maximum is taken with respect to trajectories satisfying the constraints (10.1)–(10.4) with $K(0) = K$. This of course assumes that a maximum exists for $K(0) = K$. Also, if we want to index the problem by K, we really ought to index the optimal trajectories by K, but we shall not. The thing to note about the value function is that it satisfies **Bellman's Principle of Optimality**, which states that if I^*, K^* are optimal trajectories starting at

$K(0) = K_0$, then for any time $t \geq 0$,

$$V(K_o) = \int_0^t G\big(K^*(s), I^*(s)\big)e^{-rs}\,ds + e^{-rt}V\big(K^*(t)\big). \tag{10.6}$$

What this says is that when the capital stock reaches $K^*(t)$ at time t, the optimal continuation is the same as if we reset the clock to zero, and then followed the optimal trajectory for $K_0 = K^*(t)$. This implies that if an optimal trajectory K^* exists starting at K_0, then an optimal trajectory exists for every starting value $K^*(t)$. In particular, V is defined for every $K^*(t)$. Since K^* has a derivative (namely I), it is continuous, so its range is an interval. Thus V must be defined on some interval (perhaps degenerate).

The next function we define is the **Hamiltonian** (more precisely, the **current value Hamiltonian**) for the problem,

$$H(K, I, p) = G(K, I) + pI.$$

It is the sum of the gain function and a **multiplier** or **costate** variable p multiplying the function that defines $\dot{K}$. Why we do this will become apparent later. Closely related is the **maximized Hamiltonian** $\tilde{H}$, defined by

$$\tilde{H}(K, p) = \max_{I : m(K) \leq I \leq M(K)} H(K, I, p).$$

It is the optimal value function for maximizing the Hamiltonian with respect to I.

Theorem

Assumptions

Here are the assumptions that Weitzman uses. They are satisfied for many economic problems. He notes that there are weaker assumptions under which the theorem remains true, but they are less easy to understand and the proofs are less intuitive.

1. G is concave and continuously differentiable, $G(0, 0) = 0$, and $D_1G \geq 0$ and $D_2G < 0$ everywhere.

2. The functions m and M are twice continuously differentiable, m is convex and nonincreasing, and M is concave and nondecreasing (so for $K > 0$, $m'(K) \leq 0$, $m''(K) \geq 0$, $M'(K) \geq 0$, $M''(K) \leq 0$). In addition, for $K \geq 0$, we assume $m(K) \leq 0 \leq M(K)$. To

make sure that K never becomes negative, we also assume $m(0) = 0$. Even if we do not allow capital to be consumed, it may still depreciate, in which case we generally take $m(K) = -\delta K$.

3. An optimal trajectory exists for $K_0 = 0$.

4. **Accessibility Hypothesis:** Define $R(K) = -D_1G(K, 0)/D_2G(K, 0)$. If there exists $\hat{K}$ satisfying $R(\hat{K}) = r$ (that is, $\hat{K}$ is a candidate for an optimal steady state), then $R'(\hat{K}) < 0$ (which implies $\hat{K}$ is locally unique) and $m(\hat{K}) < 0 < M(\hat{K})$ (which implies that $\hat{K}$ is accessible from both sides). Note that this rules out $m(K) = 0$ for all K if such a $\hat{K}$ exists.

The (One-Dimensional) Maximum Principle

Under the assumptions above, the pair of trajectories (K^*, I^*) is optimal (within the class of piecewise continuous trajectories) if and only if there exists a trajectory p^* of the costate variable such that for all $t \geq 0$,

$$p^*(t) \geq 0; \tag{10.7}$$

at each time t, $I(t)$ is chosen to

$$\max_I H\big(K(t), I, p(t)\big) \text{ subject to } m\big(K(t)\big) \leq I \leq M\big(K(t)\big),$$

that is,

$$H\big(K^*(t), I^*(t), p^*(t)\big) = \tilde{H}\big(K^*(t), p^*(t)\big); \tag{10.8}$$

the trajectory p^* satisfies

$$\dot{p}^*(t) = -D_1\tilde{H}\big(K^*(t), p^*(t)\big) + rp^*(t); \tag{10.9}$$

*and the following **transversality condition** holds,*

$$\lim_{t\to\infty} p^*(t)K^*(t)e^{-rt} = 0. \tag{10.10}$$

Moreover, the value function V is concave, continuously differentiable, nondecreasing, non-negative, and its derivative is the costate variable. That is, for all $t \geq 0$,

$$p^*(t) = V'\big(K^*(t)\big).$$

Let me just say here that the proof proceeds by defining the **wealth function**

$$W(t) = V\big(K^*(t)\big),$$

and using Bellman's optimality principle

$$V(K(0)) = \int_0^t G\big(K^*(s), I^*(s)\big)e^{-rs}\,ds + e^{-rt}V\big(K^*(t)\big).$$

to write

$$W(t) = e^{rt}\big[W(0) - \int_0^t G\big(K^*(s), I^*(s)\big)e^{-rs}\,ds\big],$$

which proves that W is differentiable. Since V is concave (this is easy to show), the chain rule for left- and right-hand derivatives implies that V is differentiable and

$$V'\big(K^*(t)\big) = \frac{\dot{W}(t)}{I^*(t)},$$

provided $I^*(t) \neq 0$. (The case $I^*(t) = 0$ requires a bit more work.) This now allows us to define $p^*(t)$ to be $V'\big(K^*(t)\big)$, and the remaining properties follow by more or less standard methods. Since a differentiable concave function is continuously differentiable, we conclude that $p^*(t)$ is continuous.

Commentary

On p^* and the Hamiltonian

According to the theorem, the costate variable p^* is the derivative of the value function, that is, it is the marginal value of a unit of capital to the firm, or the **shadow price** of investment. It is precisely the value of investment. The Hamiltonian is the sum

$$G(K, I) + pI,$$

the sum of the net income plus the value of investment. The fact that this is maximized at each point in time says that the firm should choose its investment to maximize the sum of its dividends G plus retained earnings p^*I, where the retained capital I is valued at its true marginal value $p^* = V'(K^*)$.

On $\dot{p}^*$

By the Envelope Theorem, *if $I^*(t)$ is an interior maximizer of the Hamiltonian*, the derivative of the maximized Hamiltonian with respect to K or p is the partial derivative of the Hamiltonian. That is,

$$D_1\tilde{H}\big(K^*(t), p^*(t)\big) = D_1H\big(K^*(t), I^*(t), p^*(t)\big) = D_1G\big(K^*(t), I^*(t)\big)$$

and

$$D_2\tilde{H}\big(K^*(t), p^*(t)\big) = D_3H\big(K^*(t), I^*(t), p^*(t)\big) = I^*(t).$$

In this case, (10.9) can be rewritten as

$$\dot{p}^*(t) = -D_1G\big(K^*(t), I^*(t)\big) + rp^*(t). \tag{10.11}$$

This can be interpreted as a **no-arbitrage condition**. At time t I can buy Δ units of capital at a price $p(t)$ and use it earn an incremental flow of income at the rate $D_1G \cdot \Delta$ for a length of time ϵ, and then resell it time $t + \epsilon$ at a price $p(t + \epsilon)$. The gain from this is

$$\Delta\big[p(t + \epsilon) - p(t) + \epsilon D_1G\big].$$

Or I could take $p(t)\Delta$ and lend it at the interest rate r for a period of length ϵ and earn $p(t)\Delta\epsilon r$. Absence of arbitrage implies that these two strategies must have the same return, or

$$p(t + \epsilon) - p(t) + \epsilon D_1G = p(t)\epsilon r.$$

Dividing by ϵ and taking the limit as $\epsilon \to 0$ implies (10.11).

Stationary optima

A stationary optimum need not exist, but suppose $\hat{K} > 0$ is a stationary optimum. That is, if $K_0 = \hat{K}$, then $K^*(t) = \hat{K}$ for all $t \geq 0$. Then $I^*(t) = 0$ for all t. If this is an interior maximizer of the Hamiltonian, then the first order condition implies

$$D_2G(\hat{K}, 0) + p^*(t) = 0,$$

so that p^* must also be constant. Then (10.9) and (10.11) imply

$$-D_1 G\big(K^*(t), I^*(t)\big) + rp^*(t) = 0,$$

or in terms of the stationary return on capital function R,

$$R(\hat{K}) = r.$$

The transversality condition

To see the necessity of the transversality condition, first use the Principle of Optimality (10.6) to get

$$\begin{aligned} e^{-rt}V\big(K^*(t)\big) &= V(K_0) - \int_0^t G\big(K^*(s), I^*(s)\big)e^{-rs}\,ds \\ &= \int_t^\infty G\big(K^*(s), I^*(s)\big)e^{-rs}\,ds \end{aligned}$$

Since the integral is convergent, we have

$$\lim_{t\to\infty} e^{-rt}V\big(K^*(t)\big) = \lim_{t\to\infty} \int_t^\infty G\big(K^*(s), I^*(s)\big)e^{-rs}\,ds = 0. \tag{10.12}$$

Now we use the concavity of V. Since concave functions lie below their tangent lines (Theorem 5.5)

$$V(0) \le V(K) + V'(K)(0 - K)$$

for all K. In particular, for $K = K^*(t)$, using the fact that $p^*(t) = V'\big(K^*(t)\big)$, we can rearrange this to get

$$V\big(K^*(t)\big) - V(0) \ge p^*(t)K^*(t) \ge 0$$

for all $t > 0$. Thus

$$e^{-rt}\left(V\big(K^*(t)\big) - V(0)\right) \ge e^{-rt}p^*(t)K^*(t) \ge 0. \tag{10.13}$$

Thus (10.12) and (10.13) imply

$$\lim_{t\to\infty} e^{-rt}p^*(t)K^*(t) = 0.$$

The economics of the transversality condition

The transversality condition also has an economic interpretation as another no-arbitrage condition. Suppose it failed—that is, suppose that

$$\limsup_{t\to\infty} e^{-rt} p^*(t) K^*(t) = A > 0.$$

Suppose I adopt the strategy of running the firm until time T, then selling it and investing the proceeds at the interest rate r. The present value of this is

$$\int_0^T G\big(K^*(t), I^*(t)\big) e^{-rt}\, dt + p^*(T) K^*(T) e^{-rT}.$$

By choosing T large enough I can make this arbitrarily close to

$$\int_0^\infty G\big(K^*(t), I^*(t)\big) e^{-rt}\, dt + A,$$

creating an arbitrage profit of just less than A. In order for this not to profitable, the transversality condition must hold.

The Wealth and Income Version of the Maximum Principle

This statement is sometimes called the Hamilton–Jacobi formulation, or Jacobi's integral form of Hamilton's equations of motion. *Under the assumptions here, the feasible trajectories* (K^*, I^*) *are optimal if and only if there exists a continuous nonnegative price trajectory* p^* *satisfying for all* $t \geq 0$*,*

$$\begin{aligned} rV\big(K^*(t)\big) &= G\big(K^*(t), I^*(t)\big) + p^*(t) I^*(t) \\ &= \tilde{H}\big(K^*(t), I^*(t)\big). \end{aligned} \tag{10.14}$$

Let's interpret this in economic terms. On the left-hand side, $V\big(K^*(t)\big)$ is the value of the optimal time-t capital stock, in other words it is **market value of the firm's equity shares** (wealth). So $rV\big(K^*(t)\big)$ is flow of interest that this equity would generate if invested at the market rate of interest (income). It is equated to the right-hand side, which consists of two parts: $G\big(K^*(t), I^*(t)\big)$, the instantaneous net income, that is, the **dividends paid out**; plus $p^*(t)I^*(t)$, the value of the optimal time-t investment at prices $p^*(t)$, which is the firm's **internal shadow price of capital**.

10.5 Application to non-renewable resources

Consider the problem of finding the monopolistic price of oil over time. The rate of interest is $r > 0$, the initial stock of oil is K_0. For simplicity we shall assume this is known. Let $E(t) \geq 0$ be the amount of oil (as a flow) that is pumped and sold at time t. For simplicity we shall assume the cost of pumping is negligible. The flow of revenue Φ from selling the flow quantity E at a given time is given by

$$\Phi(E) = E^{(\theta-1)/\theta},$$

where $\theta > 1$. Note that the consumer price is $\Phi(E)/E = E^{-1/\theta}$.

The monopolist therefore seeks to choose the trajectory E to maximize

$$\int_0^\infty E(t)^{(\theta-1)/\theta} e^{-rt}\, dt$$

subject to

$$\begin{aligned} K(0) &= K_0, \\ \dot{K}(t) &= -E(t), \\ E(t) &\geq 0, \end{aligned}$$

The Hamiltonian is $G(K, I) + pI$. Since in this problem $\dot{K}(t) = -E(t)$, we see that $-E$ plays the role of I, so $I \leq 0$ and

$$G(K, I) = (-I)^{(\theta-1)/\theta}.$$

Note that K does not appear in G at all! Also note that since $\theta > 1$ and $I < 0$, G is a decreasing function of I. Moreover $\partial G^2/\partial I^2 = -\frac{\theta-1}{\theta^2}(-I)^{-(\theta+1)/\theta} < 0$ so G is a concave function of (K, I), as we need for the assumption of our version of the Maximum Principle. Rewriting everything in terms of E we get as the Hamiltonian,

$$H(K, E, p) = E^{\frac{\theta-1}{\theta}} - pE \tag{10.15}$$

We now find the maximum of the Hamiltonian with respect to E, fixing K and p. The first and second partial derivatives of the Hamiltonian with respect to E are

$$\frac{\partial H}{\partial E} = \frac{\theta - 1}{\theta} E^{-1/\theta} - p$$

and

$$\frac{\partial^2 H}{\partial E^2} = -\frac{\theta - 1}{\theta^2} E^{-\frac{\theta+1}{\theta}} < 0.$$

Thus the Hamiltonian is concave in E, and $\frac{\partial H}{\partial E} \to \infty$ as $E \to 0$. Thus the maximum with respect to E occurs at $E > 0$. The first order condition for an interior maximum with respect to E is

$$\frac{\theta - 1}{\theta} E^{-1/\theta} - p = 0,$$

or

$$E^*(K, p) = \left(\frac{\theta - 1}{p\theta}\right)^{\theta}. \tag{10.16}$$

The maximized Hamiltonian is therefore

$$\begin{aligned} \tilde{H}(K, p) &= \left(\frac{\theta - 1}{p\theta}\right)^{\theta - 1} - p\left(\frac{\theta - 1}{p\theta}\right)^{\theta} \\ &= p^{1-\theta}\left(\frac{\theta - 1}{\theta}\right)^{\theta}\left(\left(\frac{\theta - 1}{\theta}\right)^{-1} - 1\right) \qquad (10.17) \\ &= \frac{p^{1-\theta}}{\theta - 1}\left(\frac{\theta - 1}{\theta}\right)^{\theta}. \end{aligned}$$

The three necessary and sufficient optimality conditions are (substituting $-E$ for I):

$$p^*(t) \geq 0$$

$$\begin{aligned} \dot{p}^*(t) &= -D_1\tilde{H}\big(K^*(t), p^*(t)\big) + rp^*(t) \\ &= rp^*(t) \end{aligned} \tag{10.18}$$

(since K does not appear in $\tilde{H}$). Also, $E^*(t)$ maximizes the Hamiltonian, so by the above,

$$E^*(t) = \left(\frac{\theta - 1}{p^*(t)\theta}\right)^{\theta}. \tag{10.19}$$

Finally,

$$\lim_{t\to\infty} p^*(t)K^*(t)e^{-rt} = 0, \tag{10.20}$$

From the differential equation (10.18) we have that

$$p^*(t) = p^*(0)e^{rt}.$$

The trick is to figure out $p^*(0)$.

But before we do that, let's rewrite (10.19) as

$$E^*(t) = \left(\frac{\theta - 1}{p^*(t)\theta}\right)^\theta = \underbrace{\left(\frac{\theta - 1}{p^*(0)\theta}\right)^\theta}_{=E^*(0)} e^{-r\theta t} > 0. \tag{10.21}$$

That is, the extraction never stops. From (10.20)

$$p^*(t)K^*(t)e^{-rt} = p^*(0)e^{rt}K^*(t)e^{-rt} = p^*(0)K^*(t) \to 0,$$

which implies

$$K^*(t) \to 0.$$

That is, all of the oil will eventually be extracted. This gives us the leverage we need to pin down $p^*(0)$. For

$$K(t) = K_0 - \int_0^t E(\tau)\,d\tau$$

so the condition that all the oil is extracted can be written

$$\int_0^\infty E(\tau)\,d\tau = K_0.$$

From (10.21), this becomes

$$\begin{aligned} K_0 &= \int_0^\infty E^*(0)e^{-r\theta\tau}\,d\tau \\ &= E^*(0)\int_0^\infty e^{-r\theta\tau}\,d\tau \\ &= \frac{E^*(0)}{r\theta} \end{aligned}$$

or

$$E^*(0) = r\theta K_0.$$

We can use this and (10.21) to solve for $p^*(0)$:

$$p^*(0) = \frac{\theta - 1}{\theta}(r\theta K_0)^{-1/\theta} > 0.$$

Finally let $\pi(t)$ denoted the **price paid by consumers**. As remarked above

$$\pi(t) = E(t)^{-1/\theta} = E(0)e^{rt}.$$

To summarize:

$$p^*(0) = \frac{\theta - 1}{\theta}(r\theta K_0)^{-1/\theta}.$$

$$p^*(t) = p^*(0)e^{rt}.$$

$$E^*(0) = r\theta K_0.$$

$$E^*(t) = E^*(0)e^{-r\theta t}.$$

$$K^*(t) = K_0 e^{-r\theta t}.$$

$$\pi(t) = r\theta K_0 e^{rt}$$

The relevant properties are that (i) all the oil is extracted, but it takes forever; (ii) the shadow price p grows at the rate of interest over time, and this is independent of the form of the revenue but does depend on the assumption that the cost of extraction is zero and independent of the stock of oil; and (iii) the consumer price π grows at the rate of interest over time, but this is a special property of the revenue function.

Remark 10.2. Some loose ends: Note that the theorem as stated calls for an upper and lower bound on E and we only put a lower bound on E. We can take an arbitrarily large upper bound, as long as it is large enough.

We also have the technical Accessibility Hypothesis to worry about. The stationary rate of return is defined in the notes as $R(K) = -\frac{D_1 G(K,0)}{D_2 G(K,0)}$ where $G(K,E) = (-E)^{(\theta-1)/\theta}$, so $R(K) = 0$ for all K. The Accessibility Hypothesis applies if there is a $\hat{K}$ satisfying $R(\hat{K}) = r$, which never occurs since $r > 0$. Thus the Accessibility Hypothesis is vacuously satisfied.

Remark 10.3. The costate variable $p^*(t)$ acts as a shadow price the producer has to pay for extracted oil. Even though the oil is sitting there and can be freely extracted, each unit extracted reduces the stock, and so reduces the value of the stock. The costate variable p^* captures this value reduction.

Note that in this case, the ratio of the shadow price $p^*(t)$ to the consumer price $\pi(t)$ is the constant $p^*(0)/E^*(0)$. This is an artifact due the special nature of the demand curve, which has constant price elasticity θ.

Remark 10.4. How do the trajectories change as r changes? As r increases, $p^*(t)$ increases for each t. See figure 10.1. As r increases, $E^*(0)$ increases, but for large t, $E^*(t)$ decreases.

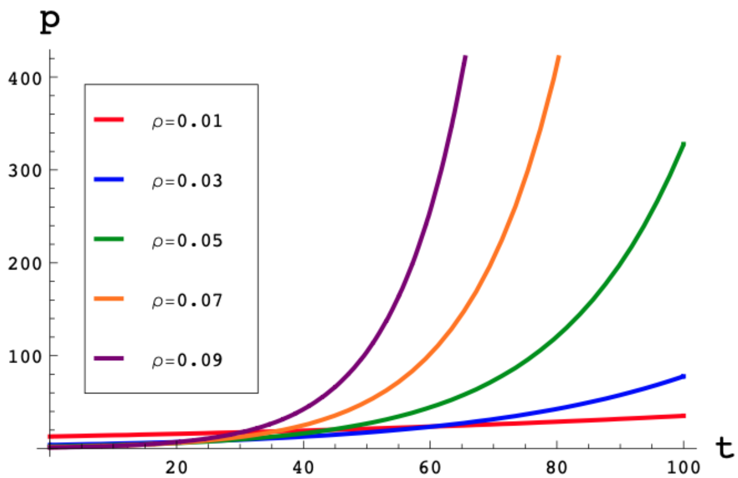


Figure 10.1: The shadow price.

See figure 10.2. As r increases, $K^*(t)$ decreases for each t. See figure 10.3. All figures are

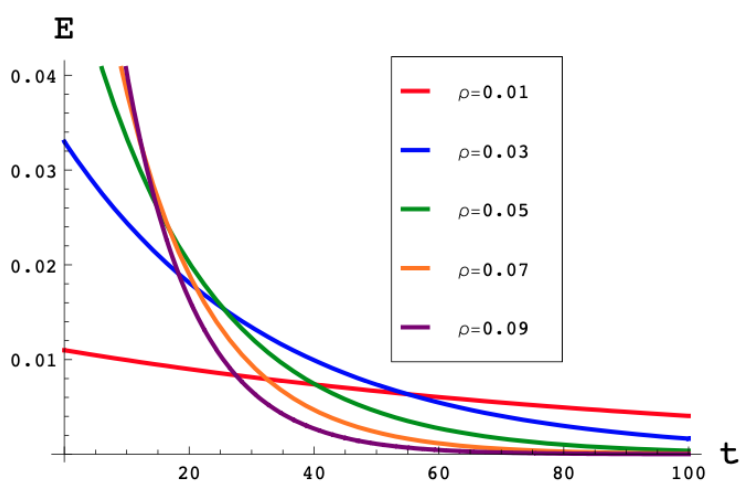


Figure 10.2: The extraction rate.

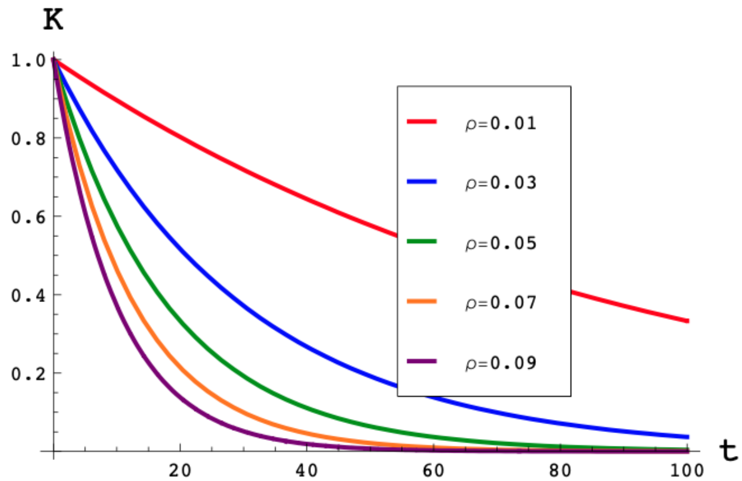


Figure 10.3: The remaining stock.

for the case $\theta = 1.1$

10.6 Appendix: The economics of first-order linear differential equations

This really has nothing to do with the maximum principle. The following theorem is a standard statement of the solution to a first order linear differential equation.

Theorem 10.5 (First order linear differential equation)**.** *Assume P, Q are continuous on the open interval I. Let $a \in I$, $b \in \mathbb{R}$.*

Then there is one and only one function $y = f(x)$ that satisfies the initial value problem

$$y' + P(x)y = Q(x) \tag{10.22}$$

with $f(a) = b$. It is given by

$$f(x) = be^{-A(x)} + e^{-A(x)} \int_a^x Q(t)e^{A(t)}\, dt$$

where

$$A(x) = \int_a^x P(t)\, dt.$$

The theorem appears a bit mysterious in this form, but I can give it an economic interpretation that makes it obvious. The first thing we will do is change the variable on which

y depends from x to time, t.

Interpret $y(t)$ as the value of a savings account at time t. At each point of time it earns an instantaneous rate of return $r(t)$. Moreover, we add a "flow" of additional savings to the account at the rate $s(t)$. Thus the rate of change of the value of the account is

$$y'(t) = r(t)y(t) + s(t). \tag{10.23}$$

Moreover, let's rewrite the initial condition as $y(t_0) = y_0$. This yields the following translation.

Theorem 10.6 (First order linear differential equation)**.** *Assume r, s are continuous on the open interval I. Let $t_0 \in I$, $y_0 \in \mathbb{R}$.*

Then there is one and only one function y that satisfies the initial value problem

$$y' = r(t)y + s(t)$$

with $y(t_0) = y_0$. It is given by

$$y(t) = [y_0 + S(t)]\, e^{\overline{r}(t)(t-t_0)}$$

where

$$\overline{r}(t) = \frac{1}{t - t_0} \int_{t_0}^{t} r(\tau)\, d\tau$$

and

$$S(t) = \int_{t_0}^{t} s(\tau) e^{-\overline{r}(\tau)(\tau - t_0)}\, d\tau.$$

But this version is obviously true!

Proof. We rely on the following well-known (easily proved) result:

$$\lim_{n\to\infty} \left(1 + (r/n)\right)^{nt} = e^{rt}.$$

That is, the result of compounding interest on a dollar continuously over t periods is e^{rt} dollars.

Case 1: $s = 0$**.** If the **instantaneous rate of return** at time t is $r(t)$, the **average rate of return** $\overline{r}(t)$ over the interval $[t_0, t]$ is just

$$\overline{r}(t) = \frac{1}{t - t_0} \int_{t_0}^{t} r(\tau)\, d\tau.$$

Now if we add nothing to the initial investment over time, that is, if $s(t) = 0$ for all t, then I claim that the value of the account at time t is given by

$$y(t) = y_0 e^{\overline{r}(t)(t-t_0)}. \tag{10.24}$$

That is, earning the varying rate of return r over the interval $[t_0, t]$ is the same as earning the average rate of return $\overline{r}$ over the interval. We can verify this by showing that y given by (10.24) solves (10.23).

$$\begin{aligned}
\frac{dy}{dt} &= \frac{d}{dt} y_0 e^{\overline{r}(t)(t-t_0)} \\
&= y_0 e^{\overline{r}(t)(t-t_0)} \frac{d}{dt} \left(\overline{r}(t)(t-t_0)\right) \\
&= y_0 e^{\overline{r}(t)(t-t_0)} \frac{d}{dt} \int_{t_0}^{t} r(\tau)\, d\tau \\
&= y_0 e^{\overline{r}(t)(t-t_0)} r(t) \\
&= r(t) y(t),
\end{aligned}$$

which is (10.23) with $s = 0$.

Case 2: General s. But in general, the additional savings $s(t)$ is not zero. In order to deal with the general case, we use the incredibly useful notion of **present value**. If you invest \$1 at time t_0 it will be worth $\$e^{\overline{r}(t)(t-t_0)}$ at time t, so the value at time t_0, that is,

the present value of \$1 at time t is $e^{-\overline{r}(t)(t-t_0)}$.

For if you invest $e^{-\overline{r}(t)(t-t_0)}$ at t_0, you will have $e^{-\overline{r}(t)(t-t_0)} e^{\overline{r}(t)(t-t_0)} = 1$ dollar at time t.

The present value of the flow $s(t)$ is $s(t)e^{-\overline{r}(t)(t-t_0)}$. The present value $S(t)$ of all the additional savings up to time t is thus

$$S(t) = \int_{t_0}^{t} s(\tau) e^{-\overline{r}(\tau)(\tau-t_0)}\, d\tau.$$

But at time t this present value will be worth an additional

$$S(t) e^{\overline{r}(t)(t-t_0)}.$$

Thus the total value of the savings account at time t is given by

$$y(t) = (y_0 + S(t))\, e^{\bar{r}(t)(t-t_0)}.$$

$\square$

Lecture 11

Introduction to Demand Theory

Read Varian, Chapters 7, 8, and 9.

11.1 Preference

Preference is comparative notion, so we represent preference as a binary relation on a set X.

$$x \succcurlyeq y \text{ means } x \text{ is as good as } y.$$
$$x \succ y \text{ means } x \text{ is better than } y.$$
$$x \sim y \text{ means } x \text{ and } y \text{ are indifferent.}$$

Formally $\succ$ and $\sim$ may be derived from $\succcurlyeq$ by

$$x \succ y \Leftrightarrow (x \succcurlyeq y \text{and} \neg\, y \succcurlyeq x)$$
$$x \sim y \Leftrightarrow (x \succcurlyeq y \text{and} y \succcurlyeq x).$$

Neoclassical economics: Assume that $\succcurlyeq$ is **regular**. That is, it has the following properties.

1. $\succcurlyeq$ is **transitive**: For all x, y, z,

$$(x \succcurlyeq y \;\&\; y \succcurlyeq z) \Rightarrow x \succcurlyeq z.$$

2. $\succcurlyeq$ is **complete**: For all x, y,

$$x \succcurlyeq y \text{ or } y \succcurlyeq x \text{ (or both).}$$

For a regular preference $\succcurlyeq$, the strict preference $\succ$ is asymmetric, transitive, and irreflexive. The indifference relation $\sim$ is symmetric, transitive, and reflexive. In other words it is an equivalence relation. The equivalence class $I(x)$ of x, that is,

$$I(x) = \{y \in X : x \sim y\}$$

is called the **indifference curve** through x. These partition the set X. For each $x, y \in X$, we have $x \in I(x)$ and

$$I(x) \cap I(y) \neq \emptyset \Rightarrow I(x) = I(y).$$

11.2 Revealed preference

Economists believe that choices reveal preferences: Choosing x when y could have been chosen reveals that $x \succcurlyeq y$. Most of us believe that choice is the only true guide to preference.

Choice functions

- X is a set of alternatives.
- A **budget** is a nonempty subset of X.
- $\mathcal{B}$ is the set of admissible budgets.
- A **choice** (or **choice function**) is a mapping c from $\mathcal{B}$ to subsets of X such that for each budget $B \in \mathcal{B}$,

$$c(B) \subset B.$$

Rational choice

A choice is rational if there are preferences for the choice to reveal. That is, choice c is **rational** if there is some binary relation $\succcurlyeq$ on X such that

$$c(B) = \{x \in B : \text{for all } y \in B,\ x \succcurlyeq y\},$$

in which case we say that $\succcurlyeq$ **rationalizes** c.

Example 11.1 (A non-rational choice)**.** $X = \{a, b, c\}$, $B_1 = \{a, b, c\}$, $B_2 = \{a, b\}$.

$$c(B_1) = \{a\}, \quad c(B_2) = \{b\}$$

is not rational, as $c(B_1) = \{a\}$ implies $a \succcurlyeq b$ and $a \succcurlyeq a$, so we must have $a \in c(B_2)$ for rationality.

Economic well-being (welfare)

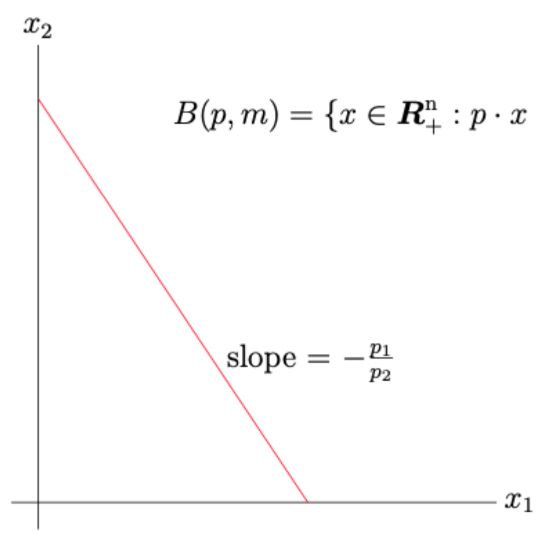
Preferences also reflect economic welfare.

A consumer would be "worse off" if forced to consume something in the budget that is not in the chosen set.

11.3 Prices and budgets

There are n commodities so $X = \mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x \geqq 0\}$. That is, alternatives are vectors of quantities of commodities. Think of them as shopping carts if you'd like.

Competitive budgets



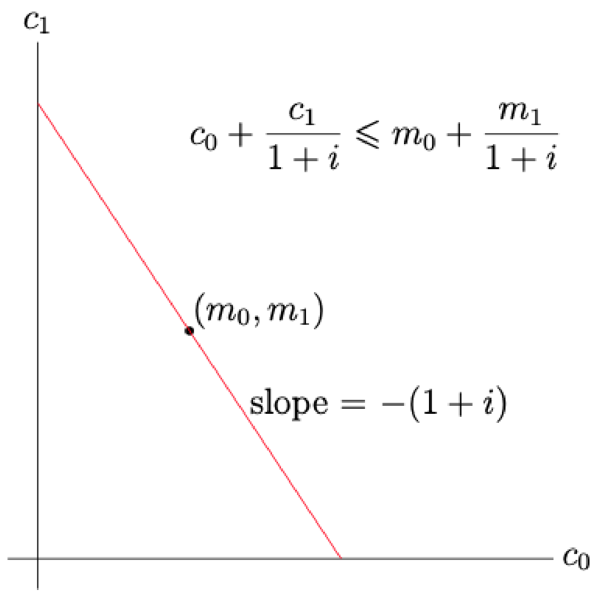
Given price vector $p \gg$ in $\mathbb{R}^n$, and income m, the budget is

$$B(p,m) = \{x \in \mathbb{R}^n_+ : p \cdot x \leq m\}.$$

Note, for $\lambda > 0$,

$$B(p,m) = B(\lambda p, \lambda m).$$

Consumption Loans budgets



Two periods $(0, 1)$, c_t is consumption at time t, m_t is real income at time t, s is savings (lending) or borrowing at $t = 0$, and i is the interest rate. The temporal budget constraint is

$$c_0 = m_0 - s$$
$$c_1 = m_1 + (1+i)s.$$

Or solving the latter for s and substituting in the former,

$$c_0 + \frac{c_1}{1+i} = m_0 + \frac{m_1}{1+i}.$$

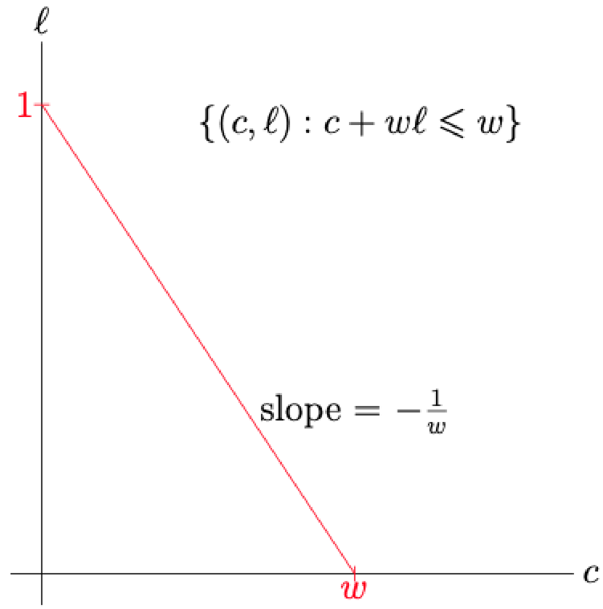
$m_0 + \frac{m_1}{1+i}$ is the **present discounted value** of income. This budget constraint is equivalent to the two separate constraints. For if

$$c_0 + \frac{c_1}{1+i} = m_0 + \frac{m_1}{1+i},$$

define $s = m_0 - c_0$. Then

$$c_0 = m_0 - s$$
$$c_1 = m_1 + (1+i)s.$$

Budget with Labor Income



w is the real wage rate. L is labor supplied. Consumption budget is

$$c \leq wL.$$

Let 1 be total amount of time in a period. Then $\ell = 1 - L$ is **leisure**. The budget becomes

$$c + w\ell \leq w.$$

11.4 Normalizing budgets

Budgets are in a sense homogeneous of degree zero. That is,

$$\text{for every } \lambda > 0, \quad B(\lambda p, \lambda m) = B(p, m).$$

Since $(p, m) \in \mathbb{R}^{n+1}$, there are only n degrees of freedom in specifying budget, and we can normmalize a (p, m) by multiplying or dividing by some $\lambda > 0$. Thus we can take as our set of budgets

$$\mathcal{B} = \{B(p, m) : p \gg 0,\ m > 0\},$$

or we could choose some good, say good n, to be the **numeraire**, and set its price to unity. In essence this measures income and prices in terms of units of good n. Since

$$B(p_1, \dots, p_n, m) = B\Big(\frac{p_1}{p_n}, \dots, \frac{p_{n-1}}{p_n}, 1, m\Big),$$

we can use as or set of budgets

$$\mathcal{B}_n = \{B(p, m) : p \gg 0,\ p_n = 1,\ m > 0\}.$$

Or we could normalize income to unity. Since

$$B(p_1, \dots, p_n, m) = B\Big(\frac{p_1}{m}, \dots, \frac{p_n}{m}, 1\Big)$$

we can use

$$\mathcal{B}_m = \{B(p, m) : p \gg 0,\ m = 1\}.$$

Or we could divide p by the sum of its components:

$$B(p_1, \dots, p_n, m) = B\Big(\frac{p_1}{\sum_{i=1}^n p_i}, \dots, \frac{p_n}{\sum_{i=1}^n p_i}, \frac{m}{\sum_{i=1}^n p_i}\Big)$$

and use

$$\mathcal{B}_{\text{sum}} = \{B(p, m) : p \gg 0,\ \sum_{i=1}^n p_i = 1,\ m > 0\}.$$

11.5 Preferences over commodity vectors

Preferences may have properties in addition to just regularity.

Definition 11.2 (Preference sets)**.**

$$P(x) = \{y \in \mathbb{R}^n_+ : y \succ x\}, \qquad U(x) = \{y \in \mathbb{R}^n_+ : y \succcurlyeq x\},$$
$$P^{-1}(x) = \{y \in \mathbb{R}^n_+ : x \succ y\}, \qquad U^{-1}(x) = \{y \in \mathbb{R}^n_+ : x \succcurlyeq y\}.$$

Continuity of preferences

Definition 11.3. $\succcurlyeq$ is **upper semicontinuous** if for each x, the upper set $U(x)$ is closed.

$\succcurlyeq$ is **lower semicontinuous** if for each x, the lower set $U^{-1}(x)$ is closed.

$\succcurlyeq$ is **continuous** if it is both upper and lower semicontinuous.

Proposition 11.4. *For a regular preference $\succcurlyeq$, the following are equivalent:*

1. *$\succcurlyeq$ is continuous.*
2. *The graph of $\succcurlyeq$, $\{(x, y) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ : x \succcurlyeq y\}$, is closed in $\mathbb{R}^n_+ \times \mathbb{R}^n_+$.*
3. *The graph of $\succ$, $\{(x, y) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ : x \succ y\}$, is open in $\mathbb{R}^n_+ \times \mathbb{R}^n_+$.*

Nonsatiation properties

Definition 11.5. $\succcurlyeq$ is **strictly monotonic** if

$$x \geqq y \;\; \& \;\; x \neq y \Rightarrow x \succ y.$$

$\succcurlyeq$ is **monotonic** if

$$x \gg y \Rightarrow x \succ y.$$

$\succcurlyeq$ is **locally nonsatiated** if for every x and every $\epsilon > 0$, there exists y satisfying

$$|y - x| < \epsilon \text{ and } y \succ x.$$

strict monotonicity $\Rightarrow$ monotonicity $\Rightarrow$ local nonsatiation

Convexity properties of preferences

Definition 11.6. $\succcurlyeq$ is **weakly convex** if

$$y \succcurlyeq x \Rightarrow (1 - \lambda)y + \lambda x \succcurlyeq x, \quad 0 < \lambda < 1.$$

$\succcurlyeq$ is **convex** if

$$y \succ x \Rightarrow (1 - \lambda)y + \lambda x \succ x, \quad 0 < \lambda < 1.$$

$\succcurlyeq$ is **strictly convex** if

$$y \succcurlyeq x \;\&\; y \neq x \Rightarrow (1 - \lambda)y + \lambda x \succ x, \quad 0 < \lambda < 1.$$

(Note: convexity does not imply weak convexity unless $\succcurlyeq$ is also upper semicontinuous.)

Proposition 11.7. *Let X be convex, and let $\succcurlyeq$ be a regular preference on X. The following are equivalent.*

1. *$\succcurlyeq$ is weakly convex.*

2. *For each x, the strict upper contour set $P(x)$ is a convex set.*

3. *For each x, the weak upper contour set $U(x)$ is a convex set.*

11.6 Utility

Definition 11.8. A function $u : X \to \mathbb{R}$ is a **utility for** $\succcurlyeq$ if

$$x \succcurlyeq y \quad \Leftrightarrow \quad u(x) \geq u(y).$$

In this case we say that u **represents** $\succcurlyeq$.

A utility is never unique. If $f : \mathbb{R} \to \mathbb{R}$ is strictly increasing, then $f \circ u$ is also a utility for $\succcurlyeq$.

Any function $u : X \to \mathbb{R}$ represents some regular preference on X.

Example 11.9 (Lexicographic preferences)**.** The **lexicographic order** on the plane is defined by

$$(x_1, x_2) \succcurlyeq (y_1, y_2) \Leftrightarrow (x_1 > y_1 \text{ or } (x_1 = y_1 \text{ and } x_2 \geq y_2))$$

Note that every indifference "curve" is a singleton!

Fact 11.10. There is no utility function that represents the lexicographic order.

11.7 Existence of utility functions

Let $X = \mathbb{R}^n_+$ and let $\succcurlyeq$ be a regular preference on X.

Proposition 11.11. *If $\succcurlyeq$ is continuous, then it can be represented by a continuous utility function on X.*

Proposition 11.12. *If $\succcurlyeq$ is upper semicontinuous, then it can be represented by an upper semicontinuous utility function on X.*

Proposition 11.13. *If $\succcurlyeq$ is continuous and strictly monotonic, then it can be represented by a strictly increasing continuous utility function on X.*

Proposition 11.14. *If $\succcurlyeq$ is weakly convex, then any utility is quasiconcave.*

If in addition, $\succcurlyeq$ is convex, then any utility is explicitly quasiconcave.

11.8 Preference Maximization

Result 11.15 (Weierstrass's Theorem)**.** If B is a nonempty, closed, bounded subset of $\mathbb{R}^n$, and $u : B \to \mathbb{R}$ is continuous, then u has a maximizer in B, that is, there exists $\bar{x} \in$ such that $u(\bar{x}) \geq u(x)$ for all $x \in B$.

Example 11.16 (Failure of a maximizer to exist)**.** Let $B = [0, 1]$ and define $u(x) = \begin{cases} x & x < 1 \\ 0 & x = 1 \end{cases}$. Then no maximizer exists. B is closed and bounded, but u is not continuous.

Let $B = (0, 1)$ and define $u(x) = x$. Then no maximizer exists. B is bounded, and u is continuous, but B is not closed.

Let $B = \mathbb{R}$ and define $u(x) = x$. Then no maximizer exists. B is closed, and u is continuous, but B is not bounded.

There is a stronger result.

Definition 11.17. A function $f : X \to \mathbb{R}$ is **upper semicontinuous** if for every $\alpha \in \mathbb{R}$, the set $\{x \in X : f(x) \geq \alpha\}$ is closed.

A function $f : X \to \mathbb{R}$ is **lower semicontinuous** if for every $\alpha \in \mathbb{R}$, the set $\{x \in X : f(x) \leq \alpha\}$ is closed.

Fact 11.18. A function $f : X \to \mathbb{R}$ is continuous if and only if it is both upper and lower semicontinuous.

Proposition 11.19. *Let B be a nonempty, closed, bounded subset of $\mathbb{R}^n$. If $u : B \to \mathbb{R}$ is upper semicontinuous, then u has a maximizer in B, that is, there exists $\bar{x} \in B$ such that $u(\bar{x}) \geq u(x)$ for all $x \in B$.*

If u is lower semicontinuous then u has a minimizer in B, that is, there exists $\underline{x} \in B$ such that $u(\underline{x}) \leq u(x)$ for all $x \in B$.

Definition 11.20. The alternative x^* is a **$\succcurlyeq$-greatest** alternative in the set B if $x^* \in B$ and for every $x \in B$, we have $x^* \succcurlyeq x$.

Proposition 11.21. *Let $B \subset \mathbb{R}^n$ be nonempty, closed and bounded, and assume that the regular preference $\succcurlyeq$ is upper semicontinuous. Then X has a $\succcurlyeq$-greatest element.*

Proposition 11.22. *Let B be convex, and assume that the regular preference $\succcurlyeq$ is strictly convex. Then a $\succcurlyeq$-greatest element is unique (if it exists).*

11.9 Demand functions

Definition 11.23 (Demand correspondence)**.**

$$x^*(p,m) = \{x \in B(p,m) : x \text{ is } \succcurlyeq\text{-greatest in } B(p,m)\}.$$

Proposition 11.24. *If $\succcurlyeq$ is continuous and $p \gg 0$, then $x^*(p,m)$ is nonempty. If $\succcurlyeq$ is strictly convex, then $x^*(p,m)$ is at most a singleton.*

Note that if $p \gg 0$, then $B(p,m)$ is closed and bounded.

Proposition 11.25. *$x^*(p,m)$ is homogeneous of degree zero in (p,m), that is,*

$$x^*(p,m) = x^*(\lambda p, \lambda m), \qquad \lambda > 0.$$

This is because $B(p,m) = B(\lambda p, \lambda m)$.

Proposition 11.26. *If $\succcurlyeq$ is locally nonsatiated, then*

$$p \cdot x^*(p,m) = m.$$

11.10 Expenditure minimization and utility maximization

Theorem 11.27. *If $\succcurlyeq$ is a continuous and locally nonsatiated regular preference on $\mathbb{R}^n_+$, and if $p \gg 0$ and $m > 0$, then*

$$x^* \textit{ maximizes } \succcurlyeq \textit{ over } B(p,m) \Leftrightarrow x^* \textit{ minimizes } p \cdot x \textit{ over } U(x^*).$$

Example 11.28 (Why $m > 0$ and/or $p \gg 0$ is needed)**.** Let $X = \mathbb{R}^2_+$. Let preferences be defined by the utility function $u(x_1, x_2) = x_1 + x_2$. (This preference relation is continuous, convex, and locally nonsatiated.) Let $x^* = (1,0)$ and $p = (0,1)$. Then x^* minimizes $p \cdot x$ over $U(x^*)$. But $B(p, p \cdot x^*) = B(p, 0)$, which is just the x_1-axis. This budget set is unbounded and no $\succcurlyeq$-greatest element exists. See figure 11.1.

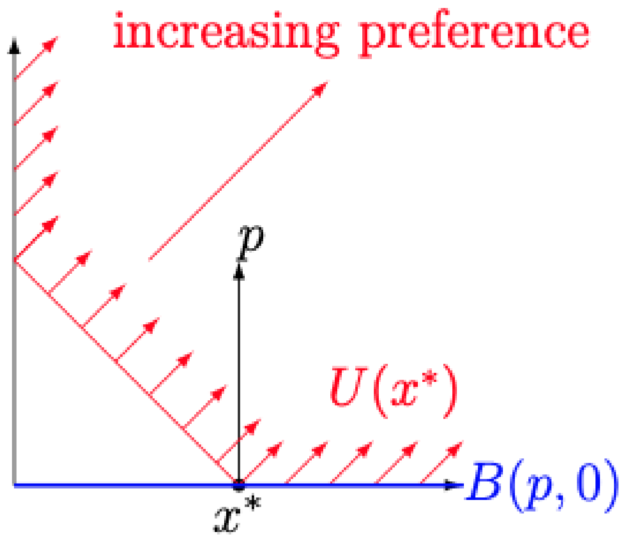


Figure 11.1: Example of nonequivalence of expenditure minimization and utility maximization.

11.11 Indirect utility and expenditure functions

Let x^* denote the ordinary demand function. The **indirect utility** v is the optimal value function for the consumer's utility maximization problem.

$$\mathrm{v}(p, m) = u\big(x^*(p, m)\big).$$

Proposition 11.29. v *is quasi-convex in* (p, m) *and homogeneous of degree zero in* (p, m).

Proof. Homogeneity follows from $B(\lambda p, \lambda m) = B(p, m)$. For quasiconvexity, we need to show that for any (p, m) and (p', m'), and $0 \leq \lambda \leq 1$ that

$$\mathrm{v}\big((1-\lambda)p + \lambda p', (1-\lambda)m + \lambda m'\big) \leq \max\{\mathrm{v}(p, m), \mathrm{v}(p', m')\}.$$

So let x_λ be demanded from the budget $B_\lambda = B\big((1-\lambda)p+\lambda p', (1-\lambda)m+\lambda m'\big)$. Observe that x_λ must belong to at least one of $B(p, m)$ or $B(p', m')$. For if this were not the case, we would have $p \cdot x_\lambda > m$ and $p' \cdot x_\lambda > m'$, which would imply that $\big((1-\lambda)p+\lambda p'\big) \cdot x_\lambda > (1-\lambda)m+\lambda m'$, contradicting the the fact that x_λ was chosen from B_λ.

Now if $x_\lambda \in B(p, m)$, by definition of the indirect utility v we would have $u(x_\lambda) \leq \mathrm{v}(p, m)$.

Ditto for (p', m'), so

$$\mathrm{v}\big((1-\lambda)p + \lambda p', (1-\lambda)m + \lambda m'\big) = u(x_\lambda) \leq \max\{\mathrm{v}(p,m), \mathrm{v}(p',m')\}.$$

$\square$

The **expenditure function** e is the optimal value function for the expenditure minimization problem

$$\min_x p \cdot x \text{ subject to } u(x) \geq \upsilon$$

The solution $\hat{x}(p, \upsilon)$ is called the **Hicksian compensated demand**.

Proposition 11.30. *$\hat{x}$ is homogeneous of degree 1 in p.*

The expenditure function is

$$e(p, \upsilon) = p \cdot \hat{x}(p, \upsilon)$$

Lecture 12

Further Topics in Demand Theory

12.1 Duality, the Envelope Theorem, and Demand

The following descriptions summarize the properties of the solutions to the utility maximization and expenditure minimization problems.

Utility Maximization Properties

- Problem statement:
$$\max_x u(x) \text{ subject to } m - p \cdot x \geq 0$$

- Optimal solution: Ordinary Walrasian Demand $x^*(p, m)$; homogeneous of degree zero in (p, m).

- Optimal value function: $v(p, m) = u(x^*(p, m))$; quasiconvex in (p, m) and homogeneous of degree zero in (p, m).

- Lagrangian:
$$L(x, \lambda; p, m) = u(x) + \lambda(m - p \cdot x)$$

- Partials w.r.t parameters:
$$\frac{\partial L(x, \lambda; p, m)}{\partial p_j} = -\lambda x_j$$
$$\frac{\partial L(x, \lambda; p, m)}{\partial m} = \lambda$$

- Envelope Theorem:

$$\frac{\partial v(p,m)}{\partial p_j} = -\lambda^*(p,m)x_j^*(p,m)$$
$$\frac{\partial v(p,m)}{m} = \lambda^*(p,m)$$

Expenditure Minimization Properties

- Problem statement:
$$\min_x p \cdot x \text{ subject to } u(x) - \upsilon \geq 0$$

- Optimal solution: Hicksian Compensated Demand $\hat{x}(p,\upsilon)$; homogeneous of degree zero in (p,m).

- Optimal value function:$e(p,\upsilon) = p \cdot \hat{x}(p,\upsilon)$; concave in p, and homogeneous of degree 1 in p.

- Lagrangian:
$$L(x,\mu;p,\upsilon) = p \cdot x - \mu\big(u(x) - \upsilon\big)$$

- Partials w.r.t parameters:

$$\frac{\partial L(x,\mu;p,\upsilon)}{\partial p_j} = x_j$$
$$\frac{\partial L(x,\mu;p,\upsilon)}{\partial \upsilon} = \mu$$

- Envelope Theorem:

$$\frac{\partial e(p,\upsilon)}{\partial p_j} = \hat{x}_j(p,\upsilon)$$
$$\frac{\partial e(p,\upsilon)}{\partial \upsilon} = \hat{\mu}(p,\upsilon)$$

Differentiating the equivalence $m = e\big(p, v(p,m)\big)$ with respect to m yields

$$1 = \frac{\partial e\big(p, v(p,m)\big)}{\partial \upsilon}\frac{\partial v(p,m)}{\partial m} = \hat{\mu}\big(p, v(p,m)\big)\lambda^*(p,m),$$

or equivalently,

$$\hat{\mu}\big(p, v(p,m)\big) = \frac{1}{\lambda^*(p,m)} \text{ and } \hat{\mu}(p,\upsilon) = \frac{1}{\lambda^*\big(p, e(p,\upsilon)\big)}.$$

From the equivalence

$$\hat{x}(p,\upsilon) = x^*\big(p, e(p,\upsilon)\big)$$

we have

$$\frac{\partial \hat{x}_i(p,\upsilon)}{\partial p_j} = \frac{\partial x_i^*\big(p, e(p,\upsilon)\big)}{\partial p_j} + \frac{\partial x_i^*\big(p, e(p,\upsilon)\big)}{\partial m}\frac{\partial e(p,\upsilon)}{\partial p_j}.$$

But $\frac{\partial e(p,\upsilon)}{\partial p_j} = \hat{x}_j(p,\upsilon) = x_j^*\big(p, e(p,\upsilon)\big)$. Set $m = e(p,\upsilon)$, and write

$$\frac{\partial \hat{x}_i(p,\upsilon)}{\partial p_j} = \frac{\partial x_i^*(p,m)}{\partial p_j} + x_j^*(p,m)\frac{\partial x_i^*(p,m)}{\partial m}$$

which implies the **Slutsky equation**

$$\frac{\partial x_i^*(p,m)}{\partial p_j} = \frac{\partial \hat{x}_i(p,\upsilon)}{\partial p_j} - x_j^*(p,m)\frac{\partial x_i^*(p,m)}{\partial m},$$

where $\upsilon = v(p,m)$, which decomposes the effect of a price change into its *substitution effect* and *income effect*. But

$$\frac{\partial \hat{x}_i(p,\upsilon)}{\partial p_j} = \frac{\partial^2 e(p,\upsilon)}{\partial p_i \partial p_j},$$

so since e is concave in p, its Hessian is negative semidefinite (and symmetric), so the matrix

$$\left[\frac{\partial x_i^*(p,m)}{\partial p_j} + x_j^*(p,m)\frac{\partial x_i^*(p,m)}{\partial m}\right] \text{ is negative semidefinite and symmetric.}$$

Consequently the diagonal terms satisfy

$$\frac{\partial \hat{x}_i(p,\upsilon)}{\partial p_i} = \frac{\partial x_i^*(p,m)}{\partial p_i} + x_i^*(p,m)\frac{\partial x_i^*(p,m)}{\partial m} \leq 0,$$

and we have the unusual **reciprocity** relation

$$\frac{\partial x_i^*(p,m)}{\partial p_j} + x_j^*(p,m)\frac{\partial x_i^*(p,m)}{\partial m} = \frac{\partial x_j^*(p,m)}{\partial p_i} + x_i^*(p,m)\frac{\partial x_j^*(p,m)}{\partial m}.$$