Chapter 8. Second-Order Conditions

8.A. Local and Global Maxima

In Chapter 7, we have discussed the sufficient conditions for optimality, confined to the context of concave programming (or more broadly, quasi-concave programming). Especially, we have concluded that when F is concave and G is convex, the first-order conditions are sufficient for maximization. More accurately, the conditions are sufficient for a *global* maximum. That is, x^* satisfying the conditions does at least as well as *any* other feasible x.

We obtain a *global* maximum in concave programming (quasi-concave programming) since the convexity (quasi-convexity) properties are defined *globally*. For example, recall the definition of convexity,

Definition 6.B.4 (Convex Function)**.** *A function $f : \mathcal{S} \to \mathbb{R}$, defined on a convex set $\mathcal{S} \subset \mathbb{R}^N$, is convex if*

$$f(\alpha x^a + (1-\alpha)x^b) \leq \alpha f(x^a) + (1-\alpha)f(x^b), \tag{6.4}$$

for all x^a, $x^b \in \mathcal{S}$ and for all $\alpha \in [0,1]$.

From the definition, (6.4) needs to hold over the full domain of f. Similar requirements appear for concavity and quasi-convextiy (quasi-concavity). These properties ensure that the desired curvature is over the full domain and thus sufficient for a *global* maximum.

The conclusions of a global maximiximum are ideal, however, in applications, we may not have functions that have the desired convexity property. In this chapter, we will focus on the curvature of the objective and constraint functions in a *small neighborhood* of the proposed optimum. The conditions are expressed in terms of the second-order derivatives of the functions at the point. Such conditions are sufficient for *local* optima – x^* satisfying the conditions does better than any other feasible x *in a sufficiently small neighborhood* of x^*.

This is a useful property when global conditions are not met. Moreover, it has a valuable by-product. It turns out that the second-order conditions play an instrumental role in determining the *comparative static* responses of the optimum choice variables x. We will discuss the comparative static result while we develop the theory of second-order conditions.

8.B. Unconstrained Maximization

We will start with the simple cases of unconstrained maximization.

First, consider the following unconstrained maximization problem with a scalar x:

$$\max_x F(x).$$

Let x^* be a candidate for the optimum choice. Expand F in a Taylor series around x^*:

$$F(x) = F(x^*) + F'(x^*)(x - x^*) + \frac{1}{2}F''(x^*)(x - x^*)^2 + ... \tag{8.1}$$

The first-order necessary condition is

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

Then (8.1) becomes

$$F(x) = F(x^*) + \frac{1}{2}F''(x^*)(x - x^*)^2 + ... \implies F(x) - F(x^*) = \frac{1}{2}F''(x^*)(x - x^*)^2 + ... \tag{8.2}$$

For x sufficiently close to x^*, the quadratic term will dominate higher-order terms in the Taylor expansion. Therefore,

(i) $F''(x^*) > 0 \implies F(x) - F(x^*) > 0 \implies F(x) > F(x^*)$.

In other words, x^* will not yield a maximum of $F(x)$ in a small neighborhood. Of course, it will not yield a maximum over the whole range of F. This argument gives a *second-order necessary* condition for x^* to yield a maximum, *local* or *global*:

$$F''(x^*) \leq 0. \tag{8.3}$$

(ii) $F''(x^*) < 0 \implies F(x) - F(x^*) < 0 \implies F(x) < F(x^*)$.

In other words, in a small neighborhood of x^*, we will have $F(x^*) > F(x)$, irrespective of the signs of higher-order terms. Thus,

$$F''(x) < 0 \tag{8.4}$$

is a *second-order sufficient* condition for x^* to yield a *local* maximum.

Note the differences between the weak inequality condition (8.3) and the strict inequality condition (8.4):

(i) (8.3) is a *necessary* condition, while (8.4) is a *sufficient* condition.

(ii) (8.3) is a condition for both *local* and *global* maximum, while (8.4) is a condition only for *local* maximum.

In the later discussions, we will focus on the local sufficiency role of second-order conditions. Please keep in mind that necessary conditions like (8.3) do exist.
A local maximum satisfying the second-order sufficient condition is called a *regular* maximum. If the maximum is "irregular", that is , if $F''(x) = 0$, then we have to look at the higher-order derivatives. Now, (8.2) becomes

$$F(x) - F(x^*) = \frac{1}{3!}F'''(x^*)(x - x^*)^3 + \frac{1}{4!}F''''(x^*)(x - x^*)^4 + ...$$

Then, $F'''(x^*) = 0$ is a necessary condition; $F'''(x^*) = 0$ and $F''''(x) < 0$ is a sufficient condition. We will leave aside such complications and focus on the *regular* maximum.

Comparative Statics. Now suppose that the problem involves a parameter θ, that is, the objective function is $F(x, \theta)$. The first-order necessary condition is

$$F_x(x^*, \theta) = 0. \tag{8.5}$$

(8.5) implicitly defines x^* as a function of θ. We want to know how the optimum choice x^* would change in response to a change of θ.

Totally differentiate (8.5), we have[1]

$$F_{xx}(x^*,\theta)\mathrm{d}x^* + F_{x\theta}(x^*,\theta)\mathrm{d}\theta = 0 \quad \text{or} \quad \frac{\mathrm{d}x^*}{\mathrm{d}\theta} = -\frac{F_{x\theta}(x^*,\theta)}{F_{xx}(x^*,\theta)}. \tag{8.6}$$

At a *regular* maximum, $F_{xx}(x^*,\theta) < 0$, the sign of $\mathrm{d}x^*/\mathrm{d}\theta$ is the same as the sign of $F_{x\theta}(x^*,\theta)$. This demonstrates how the second-order condition helps us in assessing the qualitative effects of parameter changes on the optimum choice.

An Economic Illustration. Consider the following revenue maximization problem:

$$\max_x R(x,\theta) \equiv \max_x P(x,\theta)\cdot x,$$

where x is the output and θ is a shift parameter; $P(x,\theta)$ is the inverse demand curve. Suppose

$$R_\theta(x,\theta) = P_\theta(x,\theta)\cdot x > 0$$

for all x. That is, an increase in θ shifts the demand and the revenue curves upward. By the first-order necessary condition,

$$R_x(x^*,\theta) = P_x(x^*,\theta)\cdot x^* + P(x^*,\theta) = 0. \tag{8.7}$$

Totally differentiate (8.7), we have

$$R_{xx}(x^*,\theta)\mathrm{d}x^* + R_{x\theta}(x^*,\theta)\mathrm{d}\theta = 0 \implies \frac{\mathrm{d}x^*}{\mathrm{d}\theta} = -\frac{R_{x\theta}(x^*,\theta)}{R_{xx}(x^*,\theta)} \tag{8.8}$$

At a *regular* maximum, we have $R_{xx}(x^*,\theta) < 0$. Therefore, the sign of $\mathrm{d}x^*/\mathrm{d}\theta$ is the same as the sign of $R_{x\theta}(x^*,\theta)$. Thus, an increase in θ will increase the revenue-maximizing output x^* if $R_{x\theta}(x^*,\theta) > 0$. This is true if the increase in θ shifts the *marginal* revenue upward:

$$\frac{\mathrm{d}R_x(x,\theta)}{\mathrm{d}\theta} > 0.$$

Of course, it is perfectly possible that as θ increases, the average revenue shifts up: $P_\theta(x,\theta) > 0$; but the marginal revenue shifts down: $\mathrm{d}R_x(x,\theta)/\mathrm{d}\theta < 0$. What is needed

[1] We could also write x^* as a function of θ, and differentiate (8.5) with respect to θ. By chain rule, $F_{xx}(x^*(\theta),\theta)\frac{\mathrm{d}x^*}{\mathrm{d}\theta} + F_{x\theta}(x^*(\theta),\theta) = 0 \implies \frac{\mathrm{d}x^*}{\mathrm{d}\theta} = -\frac{F_{x\theta}(x^*,\theta)}{F_{xx}(x^*,\theta)}$.

is a twist that reduces the elasticity of demand ($E_d > 0$). To see this,

$$R_x(x,\theta) = P_x(x,\theta) + P(x,\theta) = P(x,\theta)\left[P_x(x,\theta)\frac{x}{P(x,\theta)} + 1\right] = P(x,\theta)\left[1 - \frac{1}{E_d}\right].$$

Even though, an increase in θ increases $P(x,\theta)$, it is still possible that $R_x(x,\theta)$ goes down when E_d decreases with θ. If the marginal revenue does shift down ($\mathrm{d}R_x(x,\theta)/\mathrm{d}\theta < 0$), then by (8.8), a favorable shift of demand will cause output to fall.

More Choice Variables. Let us turn to the case with a vector of choice variables. Now the Taylor expansion becomes

$$\begin{aligned} F(x) &= F(x^*) + F_x(x^*)(x - x^*) + \frac{1}{2}(x - x^*)^T F_{xx}(x^*)(x - x^*) + \ldots \quad (8.9) \\ &= F(x^*) + \sum_{j=1}^{n}\left[F_j(x^*)(x_j - x_j^*)\right] + \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} F_{jk}(x_j - x_j^*)(x_k - x_k^*) + \ldots \end{aligned}$$

Here, F_{xx} is the symmetric square matrix of the second-order partial derivatives $F_{jk} \equiv \partial F/\partial x_j \partial x_k$ and the superscript T denotes the transpose operation to change the column vector into a row vector.

Similar to the scalar case, we could obtain the second-order necessary condition as well as the second-order sufficient condition. The first-order necessary condition is $F_x(x^*) = 0$. Then (8.9) becomes

$$\begin{aligned} F(x) &= F(x^*) + \frac{1}{2}(x - x^*)^T F_{xx}(x^*)(x - x^*) + \ldots \\ \Longrightarrow F(x) - F(x^*) &= \frac{1}{2}(x - x^*)^T F_{xx}(x^*)(x - x^*) + \ldots \end{aligned}$$

For x sufficiently close to x^*, the quadratic term dominates high-order terms. Therefore,

(i) $(x - x^*)^T F_{xx}(x^*)(x - x^*) \leq 0$ is the second-order necessary condition for x^* to yield a local or global maximum;

(ii) $(x - x^*)^T F_{xx}(x^*)(x - x^*) < 0$ is the second-order sufficient condition for x^* to yield a local maximum.

We will next link the second-order derivative test with the mathematical concepts of Negative (Semi-)Definiteness of matrices.

Negative (Semi-)Definite Matrix.

Definition 8.B.1 (Negative Definite)**.** *A symmetric $N \times N$ matrix M is negative definite if*

$$y^T M y < 0 \tag{8.10}$$

for all non-zero $y \in \mathbb{R}^N$.

Definition 8.B.2 (Negative Semi-definite)**.** *A symmetric $N \times N$ matrix M is negative semi-definite if*

$$y^T M y \leq 0 \tag{8.11}$$

for all $y \in \mathbb{R}^N$.

Example 1. $M = \begin{bmatrix} -2 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -2 \end{bmatrix}$ *is negative definite since for any non-zero $y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$, we have*

$$y^T M y = \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} \begin{bmatrix} -2 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} -2y_1 + y_2 & y_1 - 2y_2 + y_3 & y_2 - 2y_3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$$= -\left[y_1^2 + (y_1 - y_2)^2 + (y_2 - y_3)^2 + y_3^2 \right] < 0.$$

This result is the negative of sum of squares, and therefore non-positive. Furthermore, the result is zero only if $y_1 = y_2 = y_3 = 0$ that is, when y is the zero vector. Therefore, for any non-zero vector y, the result is always negative.

Example 2. $M = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$ *is negative semi-definite since for any $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, we have*

$$y^T M y = \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -y_1 + y_2 & y_1 - y_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = -(y_1 + y_2)^2 \leq 0.$$

This result is the negative of sum of squares, and therefore non-positive. When $y_1 = -y_2$, for example $y = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, the result is 0.

Note that a matrix M with all negative entries may not be negative definite. Example 3 illustrates the case where all entries in M is negative whereas M is not negative definite.

Example 3. $M = \begin{bmatrix} -1 & -2 \\ -2 & -1 \end{bmatrix}$ *is not negative definite since for* $y = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ *we have*

$$y^T M y = \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} -1 & -2 \\ -2 & -1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = 2 > 0.$$

Similarly, we could define positive (semi-)definite matrices analogously.

Definition 8.B.3 (Positive Definite)**.** *A symmetric $N \times N$ matrix M is positive definite if*

$$y^T M y > 0 \tag{8.12}$$

for all non-zero $y \in \mathbb{R}^N$.

Definition 8.B.4 (Positive Semi-definite)**.** *A symmetric $N \times N$ matrix M is positive semi-definite if*

$$y^T M y \geq 0 \tag{8.13}$$

for all $y \in \mathbb{R}^N$.

Remark. *A matrix that is not positive semi-definite and not negative semi-definite is called **indefinite**.*

There are various ways to check the definiteness of matrices. In Examples 1, 2 and 3, we have used the definition to check the definiteness. Below, we will introduce the determinantal test for definiteness.

Before discussing the general theorem, we need to learn some new concepts.

Definition 8.B.5 (Principal Submatrix and Principal Minor)**.** *Let M be a $N \times N$ matrix. A $k \times k$ submatrix of M formed by deleting $n-k$ rows and the same $n-k$ columns of M is called the k^{th} order **principal submatrix** of M. The determinant of a principal submatrix is called the k^{th} order **principal minor** of M.*

Example 4. *For a general* 3×3 *matrix* $M = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$.

1. *There is one* 3^{rd} *order principal minor, namely,* $\det M$;

2. *There are three* 2^{nd} *order principal minors, namely,*

   a) $\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, *formed by deleting the* 3^{rd} *row and the* 3^{rd} *column;*

   b) $\det \begin{bmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{bmatrix}$, *formed by deleting the* 2^{nd} *row and the* 2^{nd} *column;*

   c) $\det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}$, *formed by deleting the* 1^{st} *row and the* 1^{st} *column.*

3. *There are three* 1^{st} *order principal minors, namely,*

   a) $\det \begin{bmatrix} a_{11} \end{bmatrix}$, *formed by deleting the* 2^{nd} *and* 3^{rd} *rows and colomns;*

   b) $\det \begin{bmatrix} a_{22} \end{bmatrix}$, *formed by deleting the* 1^{st} *and* 3^{rd} *rows and colomns;*

   c) $\det \begin{bmatrix} a_{33} \end{bmatrix}$, *formed by deleting the* 1^{st} *and* 2^{nd} *rows and colomns.*

Definition 8.B.6 (Leading Principal Submatrix and Leading Principal Minor)**.** *Let* M *be a* $N \times N$ *matrix. The* k^{th} *order principal submatrix of* M *obtained by deleting the last* $n-k$ *rows and columns of* M *is called the* k^{th} *order* ***leading principal submatrix*** *of* M*; and its determinant is called the* k^{th} *order* ***leading principal minor*** *of* M.

Example 5. *For the general* 3×3 *matrix in Example 4,*

1. *The* 3^{rd} *order leading principal minor is* $\det M$;

2. *The* 2^{nd} *order leading principal minor is* $\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$;

3. *The* 1^{st} *order leading principal minor is* $\det \begin{bmatrix} a_{11} \end{bmatrix}$.

The following two theorems provide the algorithm for testing the definiteness of a symmetric matrix.

Theorem 8.1. *Let M be an $N \times N$ symmetric matrix. Then*

1. *M is positive definite if and only if all its leading principal minors are positive;*

2. *M is negative definite if and only if all its leading principal minors of odd order are negative; and all its leading principal minors of even order are positive.*

Theorem 8.2. *Let M be an $N \times N$ symmetric matrix. Then*

1. *M is positive semi-definite if and only if all its principal minors are non-negative;*

2. *M is negative semi-definite if and only if all its principal minors of odd order are non-positive ; and all its principal minors of even order are non-negative.*

Remark. *Please note that to check the semi-definiteness of matrices, we must unfortunately check not only the leading principal minors, but all principal minors.*

Returning to our maximization problem. We could rewrite the second-order sufficient (necessary) conditions using the terminology of (semi-)definiteness of matrices.

(i) The second-order necessary condition is that $F_{xx}(x^*)$ is negative semi-definite;

(ii) The second-order sufficient condition is that $F_{xx}(x^*)$ is negative definite.

Remark. *The second-order partial derivative matrix, F_{xx}, is called Hessian Matrix.*

Concavity. After obtaining the second-order necessary condition and sufficient condition, we would like to compare and contrast them with the property of concavity, which is defined globally.

Recall the property of concavity, expressed in terms of the first-order derivatives:

Proposition 7.A.1 (Concave Function). *A differentiable function $f : \mathcal{S} \to \mathbb{R}$, defined on a convex set $\mathcal{S} \subset \mathbb{R}^N$, is concave if and only if*

$$f_x(x^a)(x^b - x^a) \geq f(x^b) - f(x^a), \tag{7.1}$$

for all $x^a, x^b \in \mathcal{S}$.

For twice continusouly differentiable functions, this concavity property could be interpreted in terms of second-order derivatives.

Theorem 8.3. *The (twice continuously differentiable) function $f : \mathcal{S} \to \mathbb{R}$ is concave if and only if f_{xx} is negative semi-definite for every $x \in \mathcal{S}$. If f_{xx} is negative definite for every $x \in \mathcal{S}$, then the function is strictly concave.*

Proof. See Appendix A.

The link between *concavity* and the second-order necessary condition is clear: concavity requires F_{xx} to be negative semi-definite for every x, whereas the second-order necessary condition only requires F_{xx} to be negative semi-definite for the candidate optimum x^*. This is why the second-order conditions are useful: it is applicable to the functions that do not have the desired concavity property over their whole domain of definition. Of course, on the other hand, if the function do have the concavity property, it will satisfy the second-order necessary condition.

The Remark below summarizes this observation.

Remark. *To apply the second-order conditions we derived in this chapter, the objective function need not be concave (defined globally). It only needs to be "concave" at the point x^*: $F_{xx}(x^*)$ is negative semi-definite.*

Comparative Statics. Similar to the scalar variable case, we could derive the comparative static result by totally differentiating the first-order necessary condition and then applying the second-order conditions. See Example 8.4 Part I for an application.

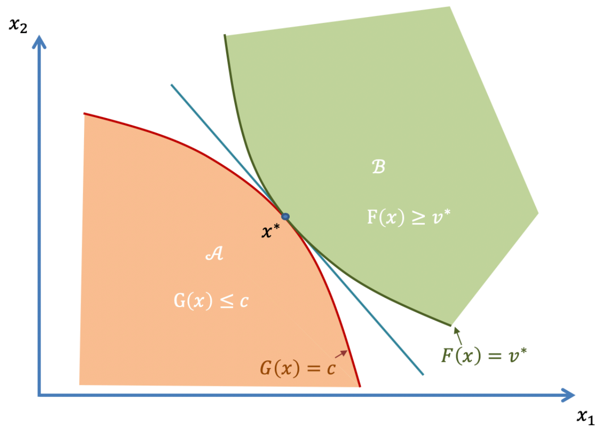
8.C. Constrained Optimization

We will begin with the simplest case of two choice variables and one equality constraint.

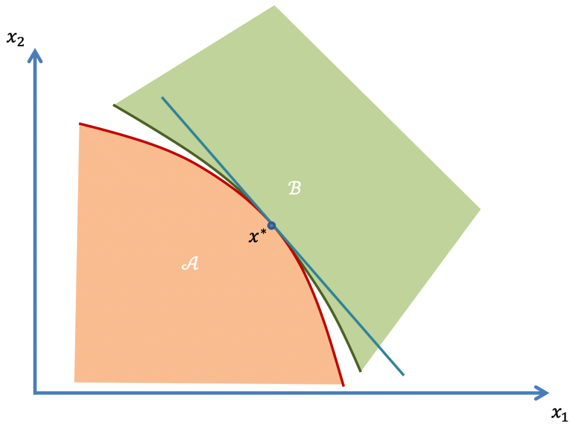
$$\max_{x_1,x_2} F(x_1, x_2)$$
$$\text{s.t. } G(x_1, x_2) = c$$

where F and G are increasing functions of their arguments.

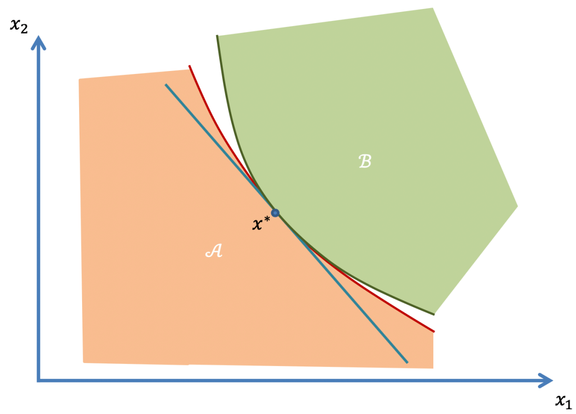
We have seen these figures (Figure 8.1) in Chapter 6 and mentioned that the *relative* curvature of F and G matters for maximization: the contour of F should be more convex than that of G.



(a)



(b)



(c)

Figure 8.1: Optimization

To express the idea algebraically, we think of x_2 as a function of x_1 along the contour of F and G, and find the second-order derivative of this function.

For F, the function of the contour is $F(x_1, x_2) = v$. Total differentiation gives

$$F_1(x_1, x_2)\mathrm{d}x_1 + F_2(x_1, x_2)\mathrm{d}x_2 = 0 \implies \frac{\mathrm{d}x_2}{\mathrm{d}x_1} = -\frac{F_1(x_1, x_2)}{F_2(x_1, x_2)}. \tag{8.14}$$

To obtain the curvature, we need to differentiate (8.14) with respect x_1 (remember now we think of x_2 as a function of x_1):

$$\frac{\mathrm{d}^2 x_2}{\mathrm{d}x_1^2} = -\frac{\mathrm{d}}{\mathrm{d}x_1}\left[\frac{F_1(x_1, x_2(x_1))}{F_2(x_1, x_2(x_1))}\right] = -\frac{F_2\left[F_{11} + F_{12}\frac{\mathrm{d}x_2}{\mathrm{d}x_1}\right] - F_1\left[F_{21} + F_{22}\frac{\mathrm{d}x_2}{\mathrm{d}x_1}\right]}{{F_2}^2}$$

$$= -\frac{F_2\left[F_{11} - F_{12}\frac{F_1}{F_2}\right] - F_1\left[F_{21} - F_{22}\frac{F_1}{F_2}\right]}{{F_2}^2} \underbrace{=}_{F_{12}=F_{21}} -\frac{{F_2}^2F_{11} - 2F_1F_2F_{12} + {F_1}^2F_{22}}{{F_2}^3}.$$

Remark. *The symmetry of the second derivative matrix follows from the Schwarz's theorem: if F has continuous second partial derivative at a, then, $\frac{\partial^2 f(a)}{\partial x_i \partial x_j} = \frac{\partial^2 f(a)}{\partial x_j \partial x_i}$.*

A similar expression could be derived for the second-order derivative along the constraint curve:

$$\frac{\mathrm{d}^2 x_2}{\mathrm{d}x_1^2} = -\frac{{G_2}^2G_{11} - 2G_1G_2G_{12} + {G_1}^2G_{22}}{{G_2}^3}.$$

The second-order sufficient condition for x^* to be a local optimum is that $\mathrm{d}^2x_2/dx_1{}^2$ along the F contour should be greater than that along the G contour:

$$-\frac{{F_2}^2F_{11} - 2F_1F_2F_{12} + {F_1}^2F_{22}}{{F_2}^3} > -\frac{{G_2}^2G_{11} - 2G_1G_2G_{12} + {G_1}^2G_{22}}{{G_2}^3}$$
$$\underbrace{\Longrightarrow}_{\text{FOC: } F_j=\lambda G_j} -\frac{\lambda^2{G_2}^2F_{11} - 2\lambda G_1\lambda G_2F_{12} + \lambda^2{G_1}^2F_{22}}{\lambda^3{G_2}^3} > -\frac{{G_2}^2G_{11} - 2G_1G_2G_{12} + {G_1}^2G_{22}}{{G_2}^3}$$
$$\underbrace{\Longrightarrow}_{G_j>0,\lambda>0} {G_2}^2\left(F_{11} - \lambda G_{11}\right) - 2G_1G_2(F_{12} - \lambda G_{12}) + {G_1}^2(F_{22} - \lambda G_{22}) < 0,$$

evaluated at x^*. This is more neatly expressed in matrix notation:

$$\det\begin{bmatrix} 0 & -G_1 & -G_2 \\ -G_1 & F_{11} - \lambda G_{11} & F_{12} - \lambda G_{12} \\ -G_2 & F_{21} - \lambda G_{21} & F_{22} - \lambda G_{22} \end{bmatrix} > 0, \tag{8.15}$$

evaluated at x^*.

Generalization to more variables and more constraints. Next, we provide without proof the conditions for the general problem with n choice variables and m equation constraints ($m < n$). Similar to the matrix notation in (8.15), we form the partitioned matrix:

$$\begin{bmatrix} 0 & -G_x \\ -{G_x}^T & F_{xx} - \lambda G_{xx} \end{bmatrix}, \tag{8.16}$$

evaluated at x^*. The top left partition 0 is $m \times m$; the bottom right partition $F_{xx} - \lambda G_{xx}$ is $n \times n$; and G_x is $m \times n$.

Remark. *The matrix*

$$\begin{bmatrix} 0 & -G_x \\ -{G_x}^T & F_{xx} - \lambda G_{xx} \end{bmatrix}$$

is called Bordered Hessian Matrix.

To check the second-order sufficient condition, we need to look at $n - m$ of the bordered Hessian's leading principal minors. Intuitively, we can think of the m constraints as reducing the problem to one with $n - m$ free variables.[2] The smallest minor we consider consisting of the truncated first $2m + 1$ rows and columns, the next consisting of the truncated first $2m + 2$ rows and columns, and so on, with the last being the determinant of the entire bordered Hessian. A sufficient condition for a local maximum of F is that the smallest minor has the same sign as $(-1)^{m+1}$ and that the rest of the principal minors alternate in sign. The result is summarized in Theorem 8.4 below.

Theorem 8.4 (Second-order Sufficient Condition for Constrained Maximization Problem)**.** *If the last $n - m$ leading principal minors of the bordered Hessian matrix at the proposed optimum x^* is such that the smallest minor (the $(2m+1)^{th}$ minor) has the same sign as $(-1)^{m+1}$ and the rest of the principal minors alternate in sign, then x^* is the local maximum of the constrained maximization problem.*

It is easy to check that (8.15) satisfies the sufficient condition for a local maximum for the two-variable one-constraint case:

1. For the two-variable one-constraint case $(n = 2, m = 1)$, we need to look at $n-m = 1$ leading principal minors. Therefore, we only need to compute the determinant of the bordered Hessian.

2. The sign requirement for maximum is $(-1)^{m+1} = (-1)^2 > 0$.

[2] For example, the maximization problem: $\max_{x,y,z} x + y^2 + z$ subject to $x + y + z = 1$ can be reduced to $\max_{x,y} x + y^2 + (1 - x - y)$ with no constraint.

Example 6. *Consider the following maximization problem with three variables ($n = 3$) and two constraints ($m = 2$):*

$$\max_{x,y,z} F(x, y, z) \equiv z$$

$$s.t.\ G^1(x, y, z) \equiv x + y + z = 12$$

$$G^2(x, y, z) \equiv x^2 + y^2 - z = 0$$

The Lagrangian is $\mathcal{L}(x, y, z, \lambda, \mu) = z + \lambda(12 - x - y - z) + \mu(-x^2 - y^2 + z)$.

The first-order necessary conditions are

$$\partial\mathcal{L}/\partial x = -\lambda - 2\mu x = 0$$

$$\partial\mathcal{L}/\partial y = -\lambda - 2\mu y = 0$$

$$\partial\mathcal{L}/\partial z = 1 - \lambda + \mu = 0$$

$$\partial\mathcal{L}/\partial \lambda = 12 - x - y - z = 0$$

$$\partial\mathcal{L}/\partial \mu = -x^2 - y^2 + z = 0$$

The stationary points are $(x^*, y^*, z^*, \lambda, \mu) = (2, 2, 8, \frac{4}{5}, -\frac{1}{5})$ *and* $(-3, -3, 18, \frac{6}{5}, \frac{1}{5})$.

The bordered Hessian matrix is

$$\begin{bmatrix} 0 & 0 & -G^1_x & -G^1_y & -G^1_z \\ 0 & 0 & -G^2_x & -G^2_y & -G^2_z \\ -G^1_x & -G^2_x & \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ -G^1_y & -G^2_y & \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ -G^1_z & -G^2_z & \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & -2x & -2y & 1 \\ -1 & -2x & -2\mu & 0 & 0 \\ -1 & -2y & 0 & -2\mu & 0 \\ -1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

We need to check $n - m = 1$ leading principal minors, i.e., we only need to check the determinant of the bordered Hessian. For local maximum, the sign requirement is $(-1)^{m+1} = (-1)^3 < 0$.

1. *For the first proposed optimum* $(x^*, y^*, z^*, \lambda, \mu) = (2, 2, 8, \frac{4}{5}, -\frac{1}{5})$*, the determinant of the bordered Hessian is* 20*;*
2. *For the second proposed optimum* $(x^*, y^*, z^*, \lambda, \mu) = (-3, -3, 18, \frac{6}{5}, \frac{1}{5})$*, the determinant of the bordered Hessian is* -20.

Thus, the 2^{nd} *proposed optimum* $(x^*, y^*, z^*, \lambda, \mu) = (-3, -3, 18, \frac{6}{5}, \frac{1}{5})$ *is a local maximum.*

Example 7. *Consider the following maximization problem with three variables* $(n = 3)$ *and one constraint* $(m = 1)$*:*

$$\max_{x,y,z} F(x, y, z) \equiv x + y + z$$

$$s.t.\ G^1(x, y, z) \equiv x^2 + y^2 + z^2 = 3$$

The Lagrangian is $\mathcal{L}(x, y, z, \lambda) = x + y + z + \lambda(3 - x^2 - y^2 - z^2)$.

The first-order necessary conditions are

$$\partial\mathcal{L}/\partial x = 1 - 2\lambda x = 0$$

$$\partial\mathcal{L}/\partial y = 1 - 2\lambda y = 0$$

$$\partial\mathcal{L}/\partial z = 1 - 2\lambda z = 0$$

$$\partial\mathcal{L}/\partial \lambda = 3 - x^2 - y^2 - z^2 = 0$$

The stationary points are $(x^*, y^*, z^*, \lambda) = (-1, -1, -1, -\frac{1}{2})$ *and* $(1, 1, 1, \frac{1}{2})$.

The bordered Hessian matrix is

$$\begin{bmatrix} 0 & -G^1_x & -G^1_y & -G^1_z \\ -G^1_x & \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ -G^1_y & \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ -G^1_z & \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix} = \begin{bmatrix} 0 & -2x & -2y & -2z \\ -2x & -2\lambda & 0 & 0 \\ -2y & 0 & -2\lambda & 0 \\ -2z & 0 & 0 & -2\lambda \end{bmatrix}$$

We need to check $n - m = 2$ *leading principal minors, i.e., the* 3^{rd} *order and the entire bordered Hessian. For local maximum, the sign requirement is* $(-1)^{m+1} = (-1)^2 > 0$ *for the* 3^{rd} *order leading principal minor and* < 0 *for the entire bordered Hessian.*

1. *For the first proposed optimum* $(x^*, y^*, z^*, \lambda) = (-1, -1, -1, -\frac{1}{2})$*, the* 3^{rd} *order leading principal minor is* $-8 < 0$ *and the determinant of the bordered Hessian is* $-12 < 0$*;*
2. *For the second proposed optimum* $(x^*, y^*, z^*, \lambda) = (1, 1, 1, \frac{1}{2})$*, the* 3^{rd} *order leading principal minor is* $8 > 0$ *and the determinant of the bordered Hessian is* $-12 < 0$.

Thus, the 2^{nd} *proposed optimum* $(x^*, y^*, z^*, \lambda) = (1, 1, 1, \frac{1}{2})$ *is a local maximum.*

Comparative Statics. For the constrained maximization problem, we could derive the comparative static results by totally differentiating the first-order necessary condition and the constrained equations, and then applying the second-order conditions. See Example 8.4 Part II for an application.

Inequality Constraints. Finally, we consider the maximization problem

$$\max_x F(x)$$
$$\text{s.t. } G(x) \leq c.$$

After applying the Kuhn-Tucker first-order necessary conditions and solving for the stationary points, we know which constraints are binding and which are not in those candidate optima. It seems that for each stationary point, to check the second-order sufficient condition, we could treat the binding constraints as the equality constraints and simply ignore the slack constraints. The intuition is correct in general, but there is one tricky point: it is possible that the inequality constraint is binding but at the same time its corresponding Lagrange multiplier is equal to 0. These inequality constraints are *degenerate inequality constraints.*

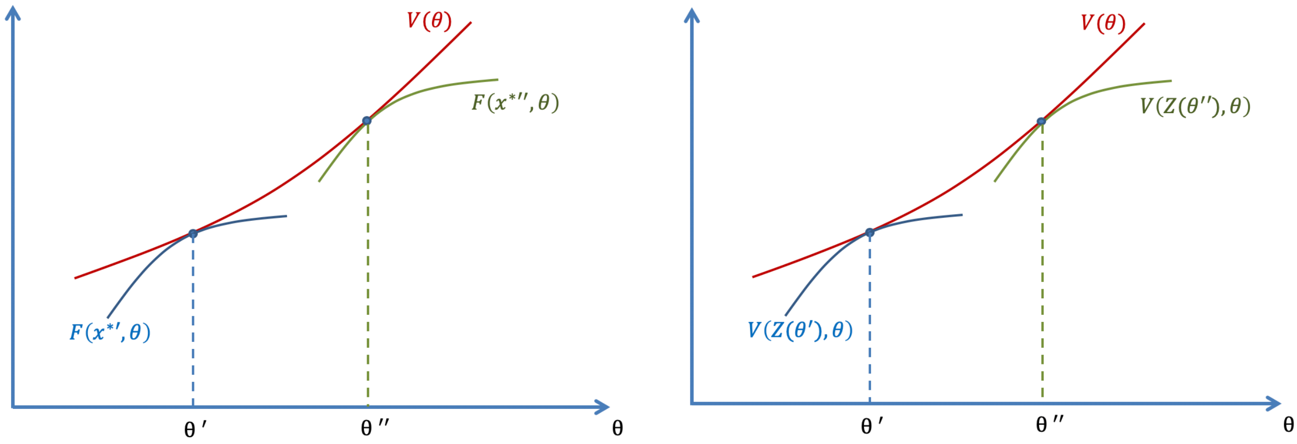
The conclusion is that to check the second-order sufficient condition, we should only keep the binding constraints with strictly positive corresponding Lagrange multipliers. In other words, we form the bordered Hessian Matrix using only the constraints with strictly positive Lagrange multipliers and then apply Theorem 8.4.

8.D. Envelope Properties

In Chapter 5, we established the envelope property of the maximum value function:

$$V(\theta) = \max_x \{F(x, \theta) \mid G(x) \leq c\}.$$

$V(\theta)$ is the upper envelope of the family of functions $F(x, \theta)$ in each of which x is held fixed. Figure 8.2a, which is the same as Figure 5.1, illustrates the envelope theorem. Subsequently, we have considered the more general problem of short-run and long-run maximum value functions, where the vector of choice variables x is partitioned into sub-vectors (y, z) and z is held fixed in the short-run. $V(\theta)$, the long-run optimum value function, is the upper envelope of the family of value functions $V(z, \theta)$, the short-run maximum value functions. Figure 8.2b, which is the same as Figure 5.4, illustrates the short-run and long-run curves.



(a) Envelope Theorem (b) Short-run and Long-run Curves

Figure 8.2: Envelope Properties

We have also mentioned the curvature properties of the envelopes. In Figure 8.2a, V is more convex than each F. In Figure 8.2b, $V(\theta)$ is more convex than $V(z, \theta)$. That is, the fewer variables are held fixed, the more convex should the maximum value function be. This second-order envelope property is the subject of this section.

Following the same notation of Chapter 5, let $Z(\theta)$ be the long-run optimum value of z. Then, the long-run and short-run value coincide at $Z(\theta)$:

$$V(\theta) = V(Z(\theta), \theta). \tag{8.17}$$

Besides, two curves are tangential at $Z(\theta)$:

$$V_\theta(\theta) = V_\theta(Z(\theta), \theta). \tag{8.18}$$

Now consider a deviation from θ to θ', we have

$$V(Z(\theta), \theta') \leq V(Z(\theta'), \theta') = V(\theta').$$

Expanding $V(Z(\theta),\theta')$ and $V(\theta')$ around θ in Taylor series, we have

$$\begin{aligned} & V(Z(\theta),\theta) + V_\theta(Z(\theta),\theta)(\theta'-\theta) + \frac{1}{2}V_{\theta\theta}(Z(\theta),\theta)(\theta'-\theta)^2 + ... \\ \le & V(\theta) + V_\theta(\theta)(\theta'-\theta) + \frac{1}{2}V_{\theta\theta}(\theta)(\theta'-\theta)^2 + ... \end{aligned} \tag{8.19}$$

By the first-order envelope properties (8.17) and (8.18), Equation (8.19) becomes

$$(V_{\theta\theta}(Z(\theta),\theta) - V_{\theta\theta}(\theta))(\theta'-\theta)^2 + ... \le 0. \tag{8.20}$$

Consider θ' sufficiently close to θ, the quadratic term in the expansion would dominate the rest of the terms. For the inequality to hold, a necessary condition is

$$V_{\theta\theta}(Z(\theta),\theta) \le V_{\theta\theta}(\theta). \tag{8.21}$$

This proves that the long-run maximum value function is at least as convex as the short-run value function at the point where the two are tangent. For suitably "regular" maxima, we have a strict inequality in (8.21).

8.E. Examples

Example 8.1: Consumer Theory.

Consider the consumer's expenditure minimization problem:

$$\begin{aligned} & \min_x px \\ & \text{s.t. } u(x) \ge u. \end{aligned} \tag{EMP}$$

In Example 5.2, we define the consumer's expenditure function $E(p,u)$ as the minimum value to the expenditure minimization problem (EMP) above. We denote the optimum quantity as the compensate demand function $C(p,u)$. The envelope property implies:

$$C(p,u) = E_p(p,u). \tag{8.22}$$

In Example 6.2, we showed that the expenditure function $E(p,u)$ is concave in p. Now by Theorem 8.3, we know that it means that $E_{pp}(p,u)$ is negative semi-definite.

Differentiating (8.22) with respect to p:

$$C_p(p, u) = E_{pp}(p, u). \tag{8.23}$$

(i) Because the second derivative matrix $E_{pp}(p, u)$ is symmetric by Schwarz's theorem, $C_p(p, u)$ is symmetric:

$$\frac{\partial C^j}{\partial p_k} = \frac{\partial C^k}{\partial p_j} = E_{jk}.$$

This is the symmetry of substitution effects of price changes.

(ii) $E_{pp}(p, u)$ is negative semi-definite. That is, $y^T E_{pp}(p, u) y \le 0$ for all $y \in \mathbb{R}^n$. In particular, we could choose $y = e^j$, where e^j is a vector with its j^{th} component equal to 1 and all other components 0. Then

$$e^{j^T} E_{pp}(p, u) e^j = E_{jj} \le 0 \implies \frac{\partial C^j}{\partial p_j} \le 0. \tag{8.24}$$

This is true for all j. Therefore, the own substitution effects of price changes are non-positive.

The second result follows even more simply from the very concept of maximum. Suppose p^a, p^b are two price vectors, and x^a, x^b are the corresponding compensated demands, both attain the same utility level u. By the definition of x^a and x^b, we have

$$p^a x^a \le p^a x^b \quad \text{and} \quad p^b x^b \le p^b x^a.$$

Adding the two inequalities

$$p^a x^a + p^b x^b \le p^a x^b + p^b x^a \implies (p^b - p^a)(x^b - x^a) \le 0. \tag{8.25}$$

(8.25) is a general version of (8.24): if p^b and p^a differ only in their j^{th} component, then (8.25) reduces to

$$(p_j^b - p_j^a)(x_j^b - x_j^a) \le 0.$$

This argument is more general in another sense: it does not require the differentiability, quasi-concavity, etc.

Example 8.2: The LeChatelier Samuelson Principal.

Consider the consumer's expenditure minimization problem (EMP) again. In this example, we focus on the second-order envelope properties. Consider a change in p_1 and compare the following two situations:

(i) The quantities of all goods are free to change optimally;

(ii) The quantity x_2 must be kept fixed at its initially optimal level.

Let $E(p_1 \mid p_{-1}, u)$ denotes the expenditure function in situation (i) and $E(p_1 \mid x_2, p_{-1}, u)$ denotes the expenditure function in situation (ii) where x_2 must be kept fixed. Let $C(p_1 \mid p_{-1}, u)$ and $C(p_1 \mid x_2, p_{-1}, u)$ be the corresponding compensated demand.

Figure 8.3 shows the envelope properties of the curves:

1. The first-order envelope property shows that the curves will be tangential at the point where x_2 is at its optimal value;
2. The second-order envelope property shows that $E(p_1 \mid p_{-1}, u)$ is more concave[3] than $E(p_1 \mid x_2^{*\prime}, p_{-1}, u)$ and $E(p_1 \mid x_2^{*\prime\prime}, p_{-1}, u)$:

$$E_{p_1p_1}(p_1 \mid p_{-1}, u) \leq E_{p_1p_1}(p_1 \mid x_2^{*\prime}, p_{-1}, u)$$
$$\text{and } E_{p_1p_1}(p_1 \mid p_{-1}, u) \leq E_{p_1p_1}(p_1 \mid x_2^{*\prime\prime}, p_{-1}, u).$$

We know from (8.23) in Example 8.1 that

$$C^1_{p_1}(p_1 \mid p_{-1}, u) = E_{p_1p_1}(p_1 \mid p_{-1}, u)$$
$$C^1_{p_1}(p_1 \mid x_2, p_{-1}, u) = E_{p_1p_1}(p_1 \mid x_2, p_{-1}, u)$$

Therefore,

$$C^1_{p_1}(p_1 \mid p_{-1}, u) \leq C^1_{p_1}(p_1 \mid x_2, p_{-1}, u)$$
$$\underbrace{\Longrightarrow}_{C^1_{p_1}(p_1 \mid p_{-1},u)\leq 0, C^1_{p_1}(p_1 \mid x_2, p_{-1},u)\leq 0} \left|C^1_{p_1}(p_1 \mid p_{-1}, u)\right| \geq \left|C^1_{p_1}(p_1 \mid x_2, p_{-1}, u)\right|$$

i.e.,

$$\left|\frac{\partial x_1}{\partial p_1}\right|_{x_2 \text{ free}} \geq \left|\frac{\partial x_1}{\partial p_1}\right|_{x_2 \text{ fixed}} \tag{8.26}$$

[3] This is a minimization problem, so the result changes from the "more convex" in the initial version to "more concave" here.

In other words, fixing quantity of some other good 2 makes the compensated demand for good 1 less responsive to its own price. Roughly speaking, any imposed rigidity in one sector of the economy causes a reduction in the responsiveness to prices in other sectors. This is true irrespective of whether good 1 and good 2 are substitutes or complements. This is known as the LeChatelier Samuelson Principle.

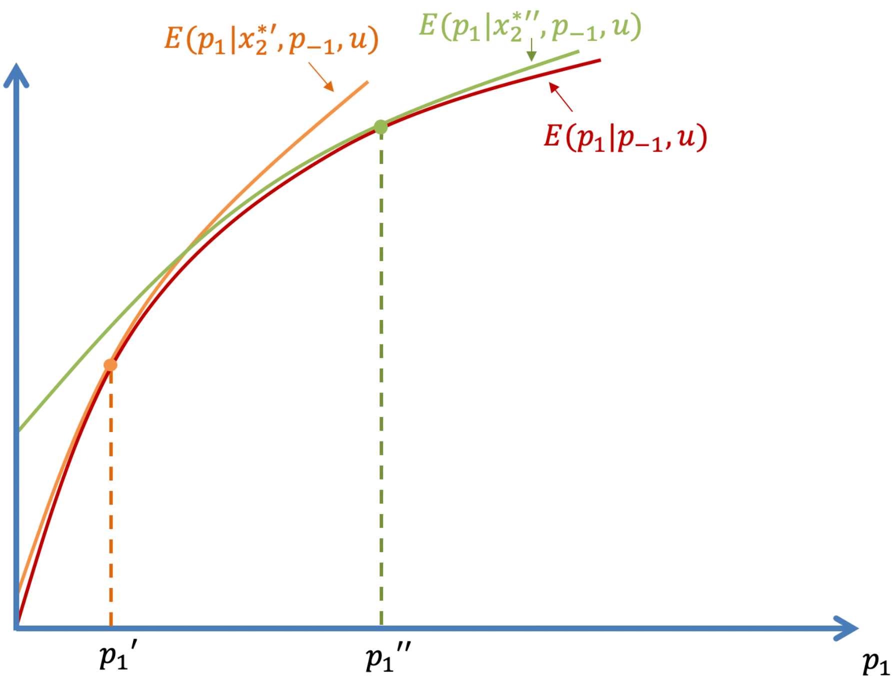


Figure 8.3: Expenditure Functions

Example 8.3: Derived Demand.

Consider the producer's cost minimization problem:

$$\min_x wx$$
$$\text{s.t. } f(x) \geq y$$

where w is the vector of input prices, x is a vector of input quantities, y is the target output level, and f is the production function. Let $C(w, y)$ be the minimized cost. Similar to the consumer's expenditure minimization problem in Example 8.1, the envelope property implies

$$x^* = C_w(w, y) \tag{8.27}$$

where x^* is the cost-minimizing input choice vector.

In the production context, different from "utility" in the consumer's context, "output" has a natural scale. Therefore, we could discuss the production technology. In particular, if the returns to scale are constant, i.e., if $f(\alpha x) = \alpha f(x)$ for all $\alpha \neq 0$, then $C(w, y)$ is proportional to output y.

To see this,

$$C(w, y) = \{w(\alpha x) \mid f(\alpha x) \geq y\} \underbrace{\iff}_{\text{constant returns to scale}} C(w, y) = \{w(\alpha x) \mid \alpha f(x) \geq y\}$$
$$\iff \frac{C(w, y)}{\alpha} = \left\{wx \mid f(x) \geq \frac{y}{\alpha}\right\} = C(w, \frac{y}{\alpha})$$

Therefore,

$$\frac{1}{\alpha}C(w, y) = C(w, \frac{y}{\alpha}), \tag{8.28}$$

for all $\alpha \neq 0$, i.e., $C(w, y)$ is proportional to y.

Let $\alpha = y$, then (8.28) becomes $\frac{1}{y}C(w, y) = C(w, 1) \implies C(w, y) = yC(w, 1)$. Therefore,

$$C(w, y) = yc(w), \tag{8.29}$$

where $c(w) = C(w, 1)$, the minimum cost of producing one unit of output.

Now consider a competitive equilibrium of an industry with a cost curve given by (8.29) and a demand curve $D(p)$. The equilibrium is attained when price equals marginal cost:

$$p = c(w). \tag{8.30}$$

The output is found from the demand curve:

$$y = D(p) \underbrace{=}_{(8.30)} D(c(w)). \tag{8.31}$$

The input demand is

$$x \underbrace{=}_{(8.27)} C_w(w, y) \underbrace{=}_{(8.29)} yc_w(w). \tag{8.32}$$

Substituting (8.31) into (8.32), we have

$$x = D(c(w))c_w(w). \tag{8.33}$$

This is called "derived demand". By chain rule,

$$\frac{\partial x_j}{\partial w_k} = D(c(w))c_{jk}(w) + D'(c(w))c_k(w)c_j(w)$$

In terms of elasticity, we have

$$\frac{w_k}{x_j}\frac{\partial x_j}{\partial w_k} = \theta_k(\sigma_{jk} - \eta), \tag{8.34}$$

where $\theta_k = \frac{w_k x_k}{yc}$ is the share of the k^{th} input in average cost; $\sigma_{jk} = \frac{cc_{jk}}{c_j c_k}$ is the elasticity of substitution between j and k; $\eta = \frac{-pD'(p)}{D(p)}$ is the elasticity of demand. Equation (8.34) splits the effect of w_k on x_j into two parts:

(i) Substitution Effect (σ_{jk}): own substitution effect is always weakly negative (concavity of c implies $c_{kk} \leq 0$, so $\sigma_{kk} \leq 0$); for a different input $j \neq k$, the effect depends on whether j and k are substitutes or complements.

(ii) Output Effect (η): an increase in w_k increases the cost, reducing the equilibrium output along the demand curve and further the demand for all inputs..

Example 8.4: Use of Second-order Conditions.

Part I. Consider a firm that buys a vector x of inputs at prices w, produced output $y = f(x)$, and sells it for revenue $R(y)$. The firm's profit maximization problem is

$$\max_x F(x, w) \equiv \max_x R(f(x)) - wx,$$

where w is a row vector of input prices.

First-order necessary condition gives

$$F_x(x^*, w) = R'(f(x^*))f_x(x^*) - w = 0. \tag{8.35}$$

Totally differentiate (8.35), we have

$$F_{xx}(x^*, w)\mathrm{d}x^* + F_{xw}(x^*, w)\mathrm{d}w^T = 0 \implies \mathrm{d}x^* = -F_{xx}(x^*, w)^{-1}F_{xw}(x^*, w)\mathrm{d}w^T. \quad (8.36)$$

From the functional form of F, we have $F_{xw}(x^*, w) = -I$. Plugging it into (8.36), we have

$$\mathrm{d}x^* = F_{xx}(x^*, w)^{-1}\mathrm{d}w^T \implies \mathrm{d}w\mathrm{d}x^* = \mathrm{d}wF_{xx}(x^*, w)^{-1}\mathrm{d}w^T.$$

By the second-order necessary condition, $F_{xx}(x^*)$ is negative semi-definite. Besides, the inverse of a negative semi-definite matrix is also negative semi-definite.[4] So

$$\mathrm{d}w\mathrm{d}x^* = \mathrm{d}wF_{xx}(x^*, w)^{-1}\mathrm{d}w^T \leq 0.$$

If the maximum is "regular", that is, the second-order sufficient condition is satisfied, then

$$\mathrm{d}w\mathrm{d}x^* < 0.$$

This result tells us how x^* would change in response to a change in w.

Part II. Consider the consumer's utility maximization problem:

$$\max_x U(x)$$
$$\text{s.t. } px = I.$$

The first-order necessary condition is

$$U_x(x^*) - \lambda p = 0. \quad (8.37)$$

We want to find the pure substitution effect of a price change. So, for the price change $\mathrm{d}p$, we compensate the consumer $\mathrm{d}I = x^{*T}\mathrm{d}p^T$. Under such compensation, the initial optimal bundle x^* is still affordable, i.e., x^* satisfies the new budget constraint:

$$(p + \mathrm{d}p)x^* = px^* + x^{*T}\mathrm{d}p^T = I + \mathrm{d}I.$$

[4] To see this, consider a symmetric negative-semidefinite matrix M and its inverse M^{-1}. First, M^{-1} is also symmetric since $(M^{-1})^T = (M^T)^{-1} \underbrace{=}_{M \text{ is symmetric}} M^{-1}$. Then, we will check $y^TM^{-1}y \leq 0$ for all y. $y^TM^{-1}y = y^TM^{-1}MM^{-1}y \underbrace{=}_{M^{-1} \text{ is symmetric}} y^T(M^{-1})^TMM^{-1}y = (M^{-1}y)^TM(M^{-1}y) \underbrace{\leq}_{M \text{ is negative semi-definite.}} 0.$

The optimal choice x^* and the Lagrange multiplier λ change as p changes. Totally differentiate (8.37) gives

$$U_{xx}(x^*)\mathrm{d}x^* - p^T\mathrm{d}\lambda - \lambda\mathrm{d}p^T = 0. \tag{8.38}$$

Totally differentiate the budget constraint gives

$$p\mathrm{d}x^* + {x^*}^T\mathrm{d}p^T = \mathrm{d}I = {x^*}^T\mathrm{d}p^T \implies p\mathrm{d}x^* = 0. \tag{8.39}$$

The second-order sufficient condition we use here should be "$\mathcal{L}_{xx}(x^*, \lambda^*)$ is negative definite". (But not the bordered Hessian is negative definite. Indeed, the bordered Hessian is not negative definite. The bordered Hessian only obeys the sign requirements in Theorem 8.4. It can be shown that if $\mathcal{L}_{xx}(x^*, \lambda^*)$ is negative definite, then the bordered Hessian obeys the sign requirements.)

The calculations are as follows. Left multiply both sides of (8.38) by $\mathrm{d}x^{*T}$ gives

$$\mathrm{d}x^{*T}U_{xx}(x^*)\mathrm{d}x^* - \mathrm{d}x^{*T}p^T\mathrm{d}\lambda - \lambda\mathrm{d}x^{*T}\mathrm{d}p^T = 0.$$

- By negative definiteness of $\mathcal{L}_{xx}(x^*, \lambda^*) = U_{xx}(x^*)$, we have $\mathrm{d}x^{*T}U_{xx}(x^*)\mathrm{d}x^* < 0$.
- By (8.39), $\mathrm{d}x^{*T}p^T = 0$.
- In addition, $\lambda > 0$.

Therefore, $\mathrm{d}p\mathrm{d}x^* < 0$.

The result indicates that the sign of the own substitution effect is negative.

Appendix A

Theorem 8.3. *The (twice continuously differentiable) function $f : \mathcal{S} \to \mathbb{R}$ is concave if and only if f_{xx} is negative semi-definite for every $x \in \mathcal{S}$. If f_{xx} is negative definite for every $x \in \mathcal{S}$, then the function is strictly concave.*

Proof.

(i) "concavity $\implies$ negative semi-definiteness": Suppose that $f(\cdot)$ is concave. Fix some $x \in \mathcal{S}$ and some arbitrary $z \in \mathbb{R}^N$, and take the second-order Taylor expansion:

$$f(x + \alpha z) = f(x) + f_x(x)(\alpha z) + \frac{\alpha^2}{2} z^T f_{xx}(x + \beta z)z, \tag{8.40}$$

for some $\beta \in [0, \alpha]$. By Proposition 7.A.1, concavity of f implies

$$f(x + \alpha z) \leq f(x) + f_x(x)(\alpha z).$$

This could be seen by taking $x^a = x, x^b = x + \alpha z$ in (7.1). Together with (8.40), we have

$$\frac{\alpha^2}{2} z^T f_{xx}(x + \beta z)z \leq 0.$$

Taking α arbitrarily small, then β is also arbitrarily small, so

$$z^T f_{xx}(x)z \leq 0,$$

which means that f_{xx} is negative semi-definite at x. Since the above is true for every $x \in \mathcal{S}$, we have proved the negative semi-definiteness of f_{xx} for every $x \in \mathcal{S}$.

(ii) "negative semi-definiteness $\implies$ concavity": Since f_{xx} is negative semidefinite for all $x \in S$,

$$z^T f_{xx}(x + \beta z)z \leq 0,$$

for all z and β. Together with (8.40), we have

$$f(x + \alpha z) \leq f(x) + f_x(x)(\alpha z)$$

for all α and z. That is,

$$f(x^b) \leq f(x^a) + f_x(x^a)(x^b - x^a)$$

for all x^a and x^b. Thus, f is concave.

(iii) "negative definiteness $\implies$ strict concavity": this part of proof is similar to (ii).

Remark. *Theorem 8.3 does not assert that negative definiteness of $f_{xx}f(x)$ must hold whenever $f(\cdot)$ is strictly concave. In fact, this is not true, $f(x) = -x^4$ is strictly concave, but* $\mathrm{d}^2 f(0)/\mathrm{d}x^2 = 0$.