Ch04. Maximum Theorem, Implicit Function Theorem and Envelope Theorem

Ping Yu

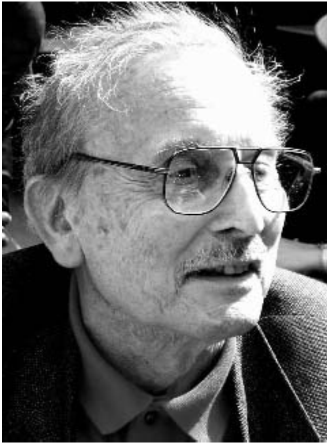
Faculty of Business and Economics
The University of Hong Kong

Overview of This Chapter

- The maximum theorem: the continuity of the optimizer and optimum with respect to (w.r.t.) a group of parameters.
- The implicit function theorem: the differentiablity of the optimizer w.r.t. a group of parameters.
- The envelope theorem: the differentiablity of the optimum w.r.t. a group of parameters.
- In the previous two chapters, we are "**moving along** a curve", that is, f and G are given (or the parameter value is given) and we are searching over $\mathbf{x}$ along f to find the optimal $\mathbf{x}^*$.
- In this chapter, we are "**shifting** a curve", i.e., we change the parameter values to shift f and G, and check how the optimizer and optimum respond to such shifting.

The Maximum Theorem

History of the Maximum Theorem



Claude J. Berge (1926-2002), French

The Problem

- Problem: We want to know how $\mathbf{x}^*$ (or the maximized function) depends on the exogenous parameters $\mathbf{a}$, rather than what $\mathbf{x}^*$ is for a particular $\mathbf{a}$.
  - For example, in consumer's problem $\max_{x_1, x_2} u(x_1, x_2)$ s.t. $p_1 x_1 + p_2 x_2 = y$, how (x_1^*, x_2^*) or $(u(x_1^*, x_2^*))$ depends on (p_1, p_2, y) rather than what (x_1^*, x_2^*) is when $p_1 = 2$, $p_2 = 7$, and $y = 25$, i.e., we are interested in what the demand function is.
- Mathematically,

$$\begin{aligned} \max_{x_1, \cdots, x_n} \; & f(x_1, \cdots, x_n, a_1, \cdots, a_k) \\ \text{s.t. } & g_1(x_1, \cdots, x_n, a_1, \cdots, a_k) = c_1, \\ & \vdots \\ & g_m(x_1, \cdots, x_n, a_1, \cdots, a_k) = c_m. \end{aligned}$$

- When is the "demand" function continuous?

Continuous Correspondence

- Define the maximized value of f when the parameters are $(a_1, \cdots, a_k)$ as

$$v(a_1, \ldots, a_k) = f(x_1^*(a_1, \ldots, a_k), \ldots, x_n^*(a_1, \ldots, a_k), a_1, \ldots, a_k).$$

- Using the feasible set, the problem can be restated as

$$\max_{x_1, \cdots, x_n} f(x_1, \cdots, x_n, a_1, \cdots, a_k)$$
$$\text{s.t. } (x_1, \cdots, x_n) \in G(a_1, \cdots, a_k),$$

  where $G(a_1, \cdots, a_k) \equiv \{(x_1, \cdots, x_n) \mid g_j(x_1, \cdots, x_n, a_1, \cdots, a_k) = c_j, \forall j\}$ is a set valued function or a correspondence.

- Two sets of vectors A and B are within ε of each other if for any vector x in one set there is a vector x' in the other set such that $x' \in B_\varepsilon(x)$.
- The correspondence G is continuous at $(a_1, \ldots, a_k)$ if $\forall\, \varepsilon > 0$, $\exists\, \delta > 0$ such that if $(a_1', \cdots, a_k')$ is within δ of $(a_1, \cdots, a_k)$ then $G(a_1', \cdots, a_k')$ is within ε of $G(a_1, \cdots, a_k)$. [Figure here]
  - If G is a function, then the continuity of G as a correspondence is **equivalent** to the continuity as a function.
  - The continuity of the functions g_j does not necessarily imply the continuity of the feasible set. (Exercise)

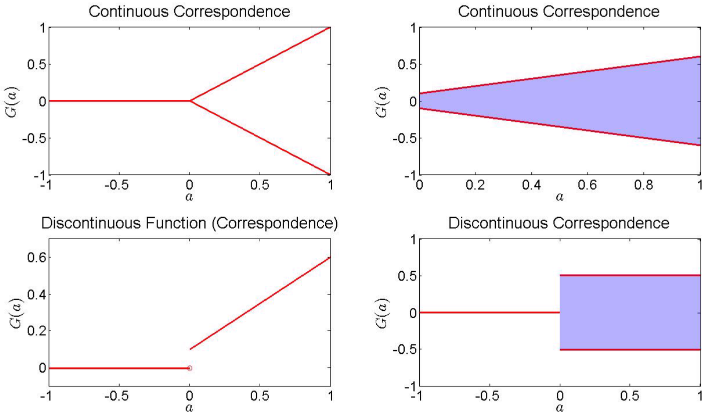


Figure: Continuous and Discontinuous Correspondence

The Maximum Theorem

Theorem

Suppose that $f(x_1, \cdots, x_n, a_1, \cdots, a_k)$ *is continuous (in* $(x_1, \cdots, x_n, a_1, \cdots, a_k)$*), that* $G(a_1, \cdots, a_k)$ *is a continuous correspondence, and that for any* $(a_1, \cdots, a_k)$ *the set* $G(a_1, \cdots, a_k)$ *is compact. Then*

(i) $v(a_1, \cdots, a_k)$ *is continuous, and*

(ii) *if* $(x_1^*(a_1, \cdots, a_k), \cdots, x_n^*(a_1, \cdots, a_k))$ *are (single valued) functions then they are also continuous.*

- Compactness of $G(a_1, \cdots, a_k)$ and continuity of f w.r.t. $\mathbf{x}$ guarantees the existence of $\mathbf{x}^*$ for any $\mathbf{a}$ by the Weierstrass Theorem.
- Uniqueness of $\mathbf{x}^*$ is guaranteed by the uniqueness theorem in Chapter 3.

The Implicit Function Theorem

History of the IFT

Augustin-Louis Cauchy (1789-1857), French

The Problem

- Problem: How to solve $\mathbf{x}^* \in \mathbb{R}^n$ from n FOCs and how sensitive of $\mathbf{x}^*$ is to the parameter?
- Suppose that we have n endogenous variables $x_1, \ldots, x_n$, m exogenous variables or parameters, $b_1, \ldots, b_m$, and n equations or equilibrium conditions

$$\begin{aligned} f_1(x_1, \cdots, x_n, b_1, \cdots, b_m) &= 0, \\ f_2(x_1, \cdots, x_n, b_1, \cdots, b_m) &= 0, \\ &\vdots \\ f_n(x_1, \cdots, x_n, b_1, \cdots, b_m) &= 0, \end{aligned}$$

  or, using vector notation,

$$\mathbf{f}(\mathbf{x}, \mathbf{b}) = \mathbf{0},$$

  where $\mathbf{f} : \mathbb{R}^{n+m} \to \mathbb{R}^n$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, and $\mathbf{0} \in \mathbb{R}^n$.
- When can we solve this system to obtain functions giving each x_i as a function of $b_1, \ldots, b_m$?

The Case of Linear Functions

- Suppose that our equations are

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n + c_{11}b_1 + \cdots + c_{1m}b_m &= 0 \\ a_{21}x_1 + \cdots + a_{2n}x_n + c_{21}b_1 + \cdots + c_{2m}b_m &= 0 \\ &\vdots \\ a_{n1}x_1 + \cdots + a_{nn}x_n + c_{n1}b_1 + \cdots + c_{nm}b_m &= 0 \end{aligned}$$

- We can write this, in matrix notation, as

$$[\mathbf{A} \mid \mathbf{C}] \begin{pmatrix} \mathbf{x} \\ \mathbf{b} \end{pmatrix} = \mathbf{0} \text{ or } \mathbf{Ax} + \mathbf{Cb} = \mathbf{0},$$

  where $\mathbf{A}$ is an $n \times n$ matrix, $\mathbf{C}$ is an $n \times m$ matrix, $\mathbf{x}$ is an $n \times 1$ (column) vector, and $\mathbf{b}$ is an $m \times 1$ vector.
- As long as $\mathbf{A}$ can be inverted or is of full rank,

$$\mathbf{x} = -\mathbf{A}^{-1}\mathbf{Cb} \text{ or } \frac{\partial \mathbf{x}}{\partial \mathbf{b}'} = -\mathbf{A}^{-1}\mathbf{C}.$$

The General Nonlinear Case

- If there are some values $(\overline{\mathbf{x}}, \overline{\mathbf{b}})$ for which $f(\overline{\mathbf{x}}, \overline{\mathbf{b}}) = \mathbf{0}$, then a parallel result holds in the **neighborhood** of $(\overline{\mathbf{x}}, \overline{\mathbf{b}})$ if we linearize f in that neighborhood.

Theorem

Suppose that $\mathbf{f} : \mathbb{R}^{n+m} \to \mathbb{R}^n$ *is a* C^1 *function on an open set* $A \subset \mathbb{R}^{n+m}$ *and that* $(\overline{\mathbf{x}}, \overline{\mathbf{b}})$ *in* A *is such that* $f(\overline{\mathbf{x}}, \overline{\mathbf{b}}) = \mathbf{0}$. *Suppose also that*

$$\frac{\partial \mathbf{f}(\overline{\mathbf{x}}, \overline{\mathbf{b}})}{\partial \mathbf{x}'} = \begin{pmatrix} \frac{\partial f_1(\overline{\mathbf{x}}, \overline{\mathbf{b}})}{\partial x_1} & \cdots & \frac{\partial f_1(\overline{\mathbf{x}}, \overline{\mathbf{b}})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n(\overline{\mathbf{x}}, \overline{\mathbf{b}})}{\partial x_1} & \cdots & \frac{\partial f_n(\overline{\mathbf{x}}, \overline{\mathbf{b}})}{\partial x_n} \end{pmatrix}$$

is of ***full rank***. *Then there are open sets* $A_1 \subset \mathbb{R}^n$ *and* $A_2 \subset \mathbb{R}^m$ *with* $\overline{\mathbf{x}} \in A_1$, $\overline{\mathbf{b}} \in A_2$ *and* $A_1 \times A_2 \subset A$ *such that for each* $\mathbf{b}$ *in* A_2 *there is exactly one* $\mathbf{g}(\mathbf{b})$ *in* A_1 *such that* $\mathbf{f}(\mathbf{g}(\mathbf{b}), \mathbf{b}) = \mathbf{0}$. *Moreover,* $\mathbf{g} : A_2 \to A_1$ *is a* C^1 *function and*

$$\left[\frac{\partial \mathbf{g}(\mathbf{b})}{\partial \mathbf{b}'}\right]_{n\times m} = -\left[\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{b}), \mathbf{b})}{\partial \mathbf{x}'}\right]_{n\times n}^{-1} \left[\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{b}), \mathbf{b})}{\partial \mathbf{b}'}\right]_{n\times m}.$$

Explanations of the IFT

- Intuition: For $(\mathbf{x},\mathbf{b})$ in a neighborhood of $(\bar{\mathbf{x}},\bar{\mathbf{b}})$ such that $\mathbf{f}(\mathbf{x},\mathbf{b})=\mathbf{0}$ with $\mathbf{x}=\mathbf{g}(\mathbf{b})$, we have
$$\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{b}),\mathbf{b})}{\partial \mathbf{x}'}\frac{\partial \mathbf{g}(\mathbf{b})}{\partial \mathbf{b}'}+\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{b}),\mathbf{b})}{\partial \mathbf{b}'}=\mathbf{0}_{n\times m},$$
so
$$\frac{\partial \mathbf{g}(\mathbf{b})}{\partial \mathbf{b}'}=-\left[\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{b}),\mathbf{b})}{\partial \mathbf{x}'}\right]^{-1}\left[\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{b}),\mathbf{b})}{\partial \mathbf{b}'}\right]\sim -\mathbf{A}^{-1}\mathbf{C} \text{ in the linear case.}$$
- This is a **local** result, rather than a **global** result.
- If $\mathbf{b}$ is not close to $\bar{\mathbf{b}}$ we may not be able to solve the system, and that for a particular value of $\mathbf{b}$ there may be many values of $\mathbf{x}$ that solve the system, but there is only one close to $\bar{\mathbf{x}}$. [Figure here]
- For all values of b close to $\bar{b}$ we can find a unique value of x close to $\bar{x}$ such that $f(x,b)=0$. However, (1) for each value of b there are other values of x far away from $\bar{x}$ that also satisfy $f(x,b)=0$, and (2) there are values of b, such as $\tilde{b}$ for which there are no values of x that satisfy $f(x,b)=0$.
- That $\frac{\partial f(\bar{x},\bar{b})}{\partial x}$ is of full rank means $|g'(\bar{x})|\neq 0$ in this example.

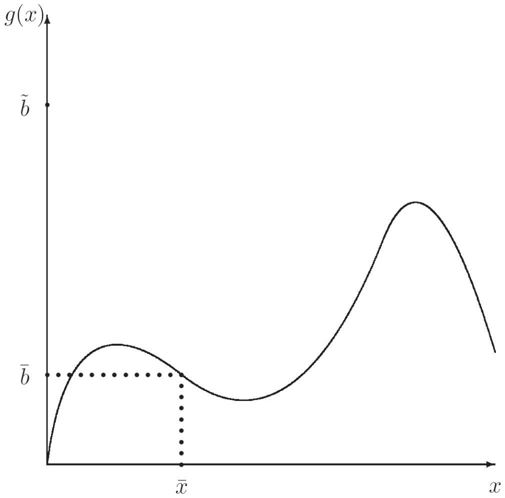


Figure: Intuition for the IFT: $f(x,b) = g(x) - b$, $x, b \in \mathbb{R}$

Comparative Statics

- The IFT does not provide conditions to guarantee the existence of $(\overline{\mathbf{x}}, \overline{\mathbf{b}})$ such that $f(\overline{\mathbf{x}}, \overline{\mathbf{b}}) = \mathbf{0}$; rather, it provides conditions such that if such an $(\overline{\mathbf{x}}, \overline{\mathbf{b}})$ exists, then we can also uniquely solve $f(\mathbf{x}, \mathbf{b}) = \mathbf{0}$ in its neighborhood.
- So the most important application of the IFT is to obtain $\frac{\partial \mathbf{g}(\mathbf{b})}{\partial \mathbf{b}'}$ rather than guarantee the existence or uniqueness of the solution.

Example

The seller of a product pays a proportional tax at a flat rate $\theta \in (0,1)$. Hence, the effective price received by the seller is $(1-\theta)P$, where P is the market price for the good. Market supply and demand are given by the differentiable functions

$$\begin{aligned} Q^d &= D(P), \text{ with } D'(\cdot) < 0 \\ Q^s &= S((1-\theta)P), \text{ with } S'(\cdot) > 0 \end{aligned}$$

and equilibrium requires market clearing, that is, $Q^s = Q^d$. Analyze, graphically and analytically, the effects of a decrease in the tax rate on the quantity transacted and the equilibrium price.

Solution to the Example

- Market clearing requires
$$S((1-\theta)P)=D(P). \tag{1}$$
This equation implicitly defines the equilibrium price as a function $P^*=P(\theta)$ of the parameter θ.
- Substituting the solution function $P(\cdot)$ back into (1), we have the identity
$$S[(1-\theta)P(\theta)]=D[P(\theta)].$$
Applying the IFT directly with $f(P,\theta)=D[P(\theta)]-S[(1-\theta)P(\theta)]$, we have
$$P'(\theta)=\frac{-PS'(\cdot)}{D'(\cdot)-(1-\theta)S'(\cdot)}=\frac{(-)}{(-)}>0.$$
- The quantity transacted in equilibrium is given by $Q^*=D[P(\theta)]$, and therefore
$$\frac{dQ^*}{d\theta}=D'(P^*)P'(\theta)<0.$$
- Graphically, a reduction in the tax rate increases the effective price received by sellers for any given market price; these are therefore willing to sell any given quantity at a lower market price. Hence the supply curve shifts down. The equilibrium price falls, and the equilibrium quantity increases. [Figure here]

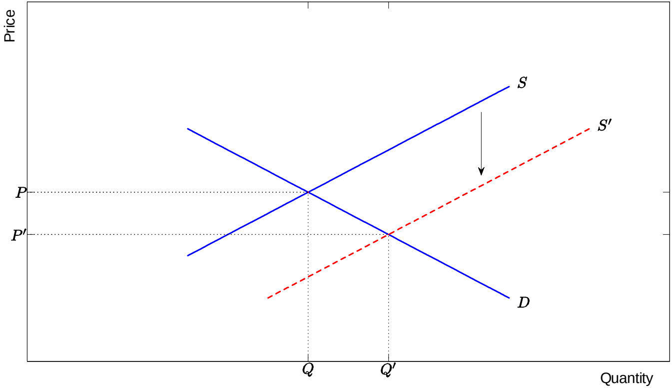


Figure: Effect of a Tax Reduction

The Envelope Theorem

Intuition for The Envelope Theorem

- Problem: How is $v(a_1, \ldots, a_k)$ sensitive to $(a_1, \ldots, a_k)$?
- Suppose we are at a maximum (in an unconstrained problem) and we change the data of the problem by a very small amount.
- Now both the solution of the problem and the value at the maximum will change.
- However at a maximum the function is flat (the first derivative is zero).
- Thus when we want to know by how much the maximized value has changed it does not matter (very much) whether or not we take account of how the maximizer changes or not.
- In the figure in the next slide, $f(x^*(a'), a') \approx f(x^*(a), a')$, i.e.,

$$\frac{\partial v(a)}{\partial a} = \underbrace{\frac{\partial f(x^*(a), a)}{\partial x}}_{=0 \text{ by FOC}} \frac{\partial x^*(a)}{\partial a} + \frac{\partial f(x^*(a), a)}{\partial a} = \frac{\partial f(x^*(a), a)}{\partial a}.$$

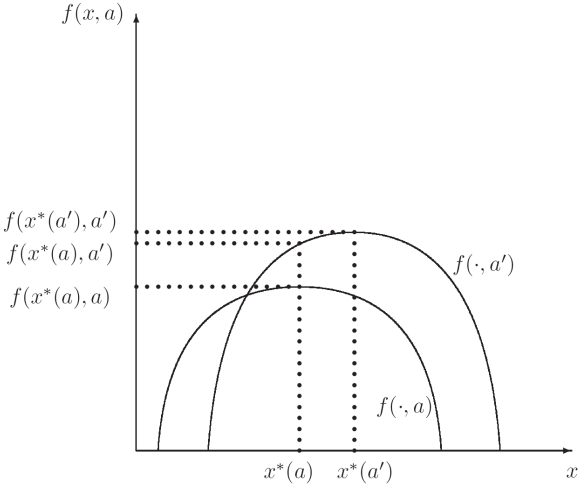


Figure: Intuition for the Envelope Theorem

SRAC and LRAC: A Motivating Example

- The short run average cost (SRAC) of producing Q units when the scale is k be given by $SRAC(Q,k)$ and the long run average cost (LRAC) of producing Q units by $LRAC(Q)$..
- Here, the scale of operation k, i.e., the size and number of plants and other fixed capital, is assumed not to be changed for some level of production Q in the short run, but will be selected to be the optimal scale for that level of production in the long run, so Q is like a and k is like x above.
- In other words,

$$LRAC(Q) = \min_k SRAC(Q,k).$$

- Let us denote, for a given value Q, the optimal level of k by $k(Q)$.
- Drawing one short run average cost curve for each of the (infinite) possible values of k. One way of thinking about the long run average cost curve is as the "bottom" or **envelope** of these short run average cost curves. [Figure here] This implies that

$$\frac{d\ LRAC(Q)}{dQ} = \frac{\partial\ SRAC(Q,k(Q))}{\partial Q}.$$

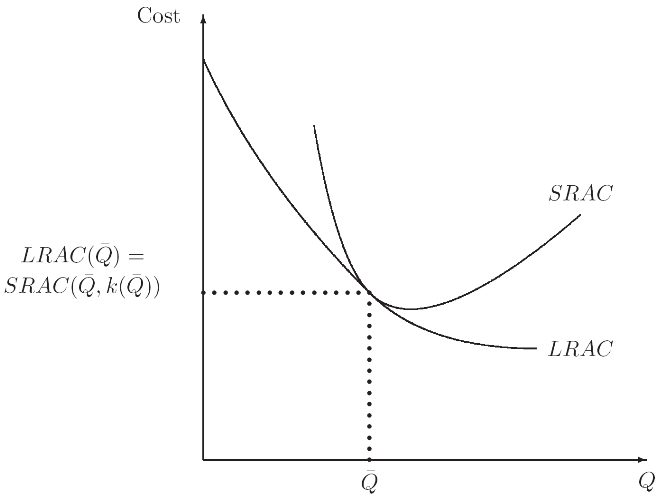


Figure: The Relationship Between SRAC and LRAC

The General Problem

- Consider again the maximization problem

$$\begin{aligned} \max_{x_1,\cdots,x_n} \; & f(x_1,\cdots,x_n,a_1,\cdots,a_k) \\ \text{s.t. } & g_1(x_1,\cdots,x_n,a_1,\cdots,a_k) = c_1, \\ & \vdots \\ & g_m(x_1,\cdots,x_n,a_1,\cdots,a_k) = c_m. \end{aligned}$$

- The Lagrangian

$$\begin{aligned} & \mathscr{L}(x_1,\cdots,x_n,\lambda_1,\cdots,\lambda_m;a_1,\cdots,a_k) \\ = \; & f(x_1,\cdots,x_n,a_1,\cdots,a_k) + \sum_{j=1}^{m} \lambda_j (c_j - g_j(x_1,\cdots,x_n,a_1,\cdots,a_k)). \end{aligned}$$

- Suppose the optimal values of x_i and λ_j are $x_i^*(a_1,\cdots,a_k)$ and $\lambda_j^*(a_1,\cdots,a_k)$, $i = 1,\cdots,n$, $j = 1,\cdots,k$. Then

$$v(a_1,\ldots,a_k) = f(x_1^*(a_1,\ldots,a_k),\ldots,x_n^*(a_1,\ldots,a_k),a_1,\ldots,a_k).$$

- The envelope theorem says that the derivative of v is equal to the derivative of $\mathscr{L}$ at the maximizing values of x and λ.

The Envelope Theorem

Theorem

If all functions are defined as above and the problem is such that the functions x^* and λ^* are well defined, then

$$\begin{aligned}
\frac{\partial v}{\partial a_h}(a_1,\cdots,a_k) =& \frac{\partial \mathscr{L}}{\partial a_h}(x_1^*(a_1,\cdots,a_k),\cdots,x_n^*(a_1,\cdots,a_k), \\
& \qquad \lambda_1^*(a_1,\cdots,a_k),\cdots,\lambda_m^*(a_1,\cdots,a_k);a_1,\cdots,a_k) \\
=& \frac{\partial f}{\partial a_h}(x_1^*(a_1,\cdots,a_k),\cdots,x_n^*(a_1,\cdots,a_k),a_1,\cdots,a_k) \\
& -\sum_{j=1}^m \lambda_j^*(a_1,\cdots,a_k)\frac{\partial g_h}{\partial a_h}(x_1^*(a_1,\cdots,a_k),\cdots,x_n^*(a_1,\cdots,a_k),a_1,\cdots,a_k)
\end{aligned}$$

for all $h=1,\cdots,k$. In matrix and vector notation,

$$\begin{aligned}
\frac{\partial v}{\partial \mathbf{a}}(\mathbf{a}) &= \frac{\partial \mathscr{L}}{\partial \mathbf{a}}(\mathbf{x}^*(\mathbf{a}),\lambda^*(\mathbf{a});\mathbf{a}) \\
&= \frac{\partial f}{\partial \mathbf{a}}(\mathbf{x}^*(\mathbf{a}),\mathbf{a}) - \frac{\partial \mathbf{g}(\mathbf{x}^*(\mathbf{a}),\mathbf{a})'}{\partial \mathbf{a}}\lambda^*(\mathbf{a}).
\end{aligned}$$

Interpretation of the Lagrange Multiplier

- In Chapter 2, we have interpreted the Lagrange multiplier as the penalty on violating the constraint.
- More rigorously, treating $c_j, j = 1, \cdots, m$, also as parameters, then by the envelope theorem,

$$\frac{\partial v}{\partial c_j}(a_1, \cdots, a_k, c_1, \cdots, c_m) = \lambda_j^*(a_1, \cdots, a_k, c_1, \cdots, c_m).$$

- Intuition: Think of f as the profit function of a firm, the g_j equation as the resource constraint, and c_j as the amount of input j available to the firm. In this situation, $\frac{\partial v}{\partial c_j}(a_1, \cdots, a_k, c_1, \cdots, c_m)$ represents the change in the optimal profit resulting from availability of one more unit of input j. Alternatively, it tells the maximum amount the firm would be willing to pay to get another unit of input j. For this reason, λ_j^* is often called the internal value, or more frequently, the shadow price of input j. It may be a more important index to the firm than the external market price of input j.

Other Applications in Microeconomics

- Consumer Theory:
  - Hotelling's Theorem
  - Hicks-Slutsky equations
  - Roy's Theorem
  - Shephard's lemma
- Production Theory:
  - Hotelling's lemma
  - Shephard's lemma (again!)