19. Constrained Optimization II

11/22/22

NB: Problems 2, 7, and 13 from Chapter 18 and problems 2 and 3 from Chapter 19 are due on Tuesday, November 29.

We continue our investigation of constrained optimization, including some the ideas surrounding the Kuhn-Tucker theory. We focus on four main areas.

1. Dependence of the optimum and optimal values on model parameters, including the Envelope Theorem – page 3.
2. Second order conditions using bordered Hessians – page 19.
3. Smoothness of Optimal Functions – page 29.
4. Constraint qualification: Necessity and alternatives – page 35.

19.1 Envelope Theorems and Lagrange Multipliers

We've used multipliers to solve optimization problems, but we haven't stopped to ask: Does the multiplier mean anything? If so, what does the multiplier mean?

That will eventually lead to questions about the relation between the optimizers, the optimized value, and any parameters of the problem. This is all summed up in the Envelope Theorem.[1]

[1] Envelope theorems describe the relation between an optimized objective and various parameters of the optimization problem. I don't know their origin, but they have long been used in economics. They can also be applied well beyond their traditional setting. E.g., see Paul Milgrom and Ilya Segal (2002), Envelope theorems for arbitrary choice sets, *Econometrica* **70**, 583–601. Paul Samuelson had shown that Le Chatilier's Principle can be derived from an Envelope theorem. See Paul Samuelson (1947), *Foundations of Economic Analysis*, Harvard Univ. Press, Cambridge, Mass.

19.1.1 Multiplier Interpretation

Let's start with the multiplier question. We first examine it in a stripped down model. Start with an objective function $f(x, y)$ and a single constraint function $h(x, y)$. That gives us the following maximization problem:

$$\begin{aligned} M(c) = \max_{(x,y)} & \ f(x, y) \\ \text{s.t.} & \ h(x, y) = c \end{aligned} \tag{19.1.1}$$

where $M(c) = f\big(x^*(c), y^*(c)\big)$ is the maximum value of f under the constraint.

The optimality condition implicitly defines the multiplier μ.

$$Df = \mu Dh.$$

Even in the simplest cases where it's useful to use a multiplier, we must have at least two goods and the first order condition is a covector equation. Let's focus on the first component.

$$\frac{\partial f}{\partial x} = \mu \frac{\partial h}{\partial x} \quad \text{or} \quad \mu = \frac{\partial f}{\partial x} \bigg/ \frac{\partial h}{\partial x}$$

The units of μ are units of f per unit of h. This happens regardless of which variable, x or y, is used to define μ. If we used y, we would get the same units in the end.

That suggests the multiplier is the marginal value of whatever it is that is being constrained, whatever h is. In this model we can also call it c. We will calculate the marginal value of c and see if it is in fact the multiplier.

19.1.2 Finding the Multiplier: A Basic Envelope Theorem

We return to the maximization problem

$$\begin{aligned} M(c) = &\max_{(x,y)} f(x,y) \\ &\text{s.t. } h(x,y) = c \end{aligned} \tag{19.1.1}$$

and regard c is an adjustable parameter. The maximizer $\big(x^*(c), y^*(c)\big)$ then depends on c. The use of $M(c)$ for the maximized value, $f\big(x^*(c), y^*(c)\big)$, emphasizes its dependence on the parameter c.

What we want to know is whether

$$\mu = M'(c) = \frac{d}{dc}\Big[f\big(x^*(c), y^*(c)\big)\Big].$$

In fact, it does.

Theorem 19.1.1. *Let f and h be $\mathcal{C}^1$ functions on $U \subset \mathbb{R}^2$. Let $\big(x^*(c), y^*(c)\big)$ be the solution to the maximization problem (19.1.1) and suppose $x^*(c)$, $y^*(c)$, and the multiplier $\mu^*(c)$ are continuously differentiable functions of c. Suppose further that NDCQ holds at $\big(x^*(c), y^*(c), \mu^*(c)\big)$. Then*

$$\mu^*(c) = M'(c) = \frac{d}{dc}\Big[f\big(x^*(c), y^*(c)\big)\Big].$$

Theorem 19.1.1 is a special case of the Envelope Theorem, which we will encounter later.

19.1.3 Envelope Theorem Assumptions

It's worth taking a moment to consider why the Envelope Theorem has the assumptions it does.

The Set $\mathcal{U}$. You might be curious why the Envelope Theorem involves a subset $\mathcal{U}$ of $\mathbb{R}^2$. Usually, we will be working with functions defined on subsets of $\mathbb{R}^2$. For example, both utility and production functions are typically defined on $\mathbb{R}^2_+$. By using a subset $\mathcal{U}$, we give the theorem enough flexibility to handle such problems.

Constraint Qualification. Constraint qualification (NDCQ) must be satisfied so that we can use the Lagrangian to find the first order conditions. See Theorems 18.11.1 18.18.1 and 18.35.1.

Differentiability. We require that the optimal choices and multipliers be $\mathcal{C}^1$ functions as we will need to take their derivatives as part of the proof. That leaves open the question of how we know these functions are $\mathcal{C}^1$, which we will address later in this chapter.

19.1.4 A Basic Envelope Theorem

Theorem 19.1.1. *Let* f *and* h *be* $\mathcal{C}^1$ *functions on* $U \subset \mathbb{R}^2$. *Let* $\big(x^*(c), y^*(c)\big)$ *be the solution to the maximization problem (19.1.1) and suppose* $x^*(c)$, $y^*(c)$, *and the multiplier* $\mu^*(c)$ *are continuously differentiable functions of* c. *Suppose further that NDCQ holds at* $\big(x^*(c), y^*(c), \mu^*(c)\big)$. *Then*

$$\mu^*(c) = M'(c) = \frac{d}{dc}\Big[f\big(x^*(c), y^*(c)\big)\Big].$$

Sketch of Proof. Since we will use the same method to prove several variations on the basic Envelope Theorem, it's worth summarizing our method upfront.

Logically, the main part of the proof starts when we use the Chain Rule to calculate the derivative of the optimal value $f\big(x^*(c), y^*(c)\big)$. We then rewrite the equation twice: Once using the first order conditions (19.1.2), and a second time using the results (19.1.3) from differentiating the constraint. At that point the result pops out.

The formal proof is arranged somewhat differently. Since we know we will need them, we pre-calculate both the first order conditions and the derivative of the constraint. Once that is done, we are ready for the main calculation, computing the derivative of the optimal value $f\big(x^*(c), y^*(c)\big)$. The pre-calculation allows the calculation to flow better, as we can plug in our previous results when needed rather than interrupting the main calculation to derive them.

19.1.5 Proof of Basic Envelope Theorem

Proof. The Lagrangian is $\mathcal{L} = f(x, y) - \mu\big(h(x, y) - c\big)$ The first order conditions are

$$0 = \frac{\partial \mathcal{L}}{\partial x} = \frac{\partial f}{\partial x} - \mu^* \frac{\partial h}{\partial x} \qquad 0 = \frac{\partial \mathcal{L}}{\partial y} = \frac{\partial f}{\partial y} - \mu^* \frac{\partial h}{\partial y}$$

so

$$\frac{\partial f}{\partial x} = \mu^* \frac{\partial h}{\partial x} \quad \text{and} \quad \frac{\partial f}{\partial y} = \mu^* \frac{\partial h}{\partial y} \tag{19.1.2}$$

with everything evaluated at $(x, y, \mu) = \big(x^*(c), y^*(c), \mu^*(c)\big)$. Differentiating the constraint $h\big(x^*(c), y^*(c)\big) = c$, we obtain

$$\frac{\partial h}{\partial x}\frac{dx^*}{dc} + \frac{\partial h}{\partial y}\frac{dy^*}{dc} = 1. \tag{19.1.3}$$

Applying the Chain Rule and using equations (19.1.2) and (19.1.3) we find

$$\begin{aligned}
M'(c) &= \frac{d}{dc}\Big[f\big(x^*(c), y^*(c)\big)\Big] \\
&= \frac{\partial f}{\partial x}\frac{dx^*}{dc} + \frac{\partial f}{\partial y}\frac{dy^*}{dc} \\
&= \left(\mu^* \frac{\partial h}{\partial x}\right)\frac{dx^*}{dc} + \left(\mu^* \frac{\partial h}{\partial y}\right)\frac{dy^*}{dc} && \text{By (19.1.2)} \\
&= \mu^* \left(\frac{\partial h}{\partial x}\frac{dx^*}{dc} + \frac{\partial h}{\partial y}\frac{dy^*}{dc}\right) \\
&= \mu^*(c). && \text{By (19.1.3)}
\end{aligned}$$

∎

19.1.6 Marginal Utility of Income

Let's apply Theorem 19.1.1 to the simple consumer's problem we examined in the previous chapter.

$$v(\mathbf{p}, m) = \max_{x} u(x_1, x_2)$$
$$\text{s.t. } p_1 x_1 + p_2 x_2 = m.$$

We treat m as the parameter. Provided the solutions are differentiable, (NDCQ holds automatically), we can apply Theorem 19.1.1 to find

$$\frac{\partial v}{\partial m}(\mathbf{p}, m) = \mu^*(m).$$

The multiplier belonging to the budget constraint is the marginal (indirect) utility of income.

19.1.7 Marginal Cost

We can also consider a simplified version of the firm's cost minimization problem.

$$c(\boldsymbol{w}, q) = \min_{\boldsymbol{z}} \boldsymbol{w} \cdot \boldsymbol{z}$$
$$\text{s.t. } f(\boldsymbol{z}) = q, .$$

We assume $Df \gg \mathbf{0}$, ensuring that NDCQ holds. If the solutions are $\mathcal{C}^1$, we have

$$\frac{\partial c}{\partial q}(\boldsymbol{w}, q) = \mu^*(q).$$

The multiplier here is the marginal cost.

Of course, we have to be careful drawing strong conclusions from these simple versions of the utility maximization and cost minimization problems. The details left out can be important. We will generalize Theorem 19.1.1 to apply to the full consumer's utility maximization problem and firm's cost minimization problem. That version can be used to establish these results for the standard consumer's utility maximization problem and firm's cost minimization problem.

19.1.8 Multiple Multipliers

Suppose we have many variables and multiple constraints, with multiple multipliers. We consider the problem

$$\begin{aligned} M(\mathbf{c}) = \max_{\mathbf{x}} & \; f(\mathbf{x}) \\ \text{s.t.} & \; h_i(\mathbf{x}) = c_i, \text{ for } i = 1, \dots, \ell \end{aligned} \tag{19.1.4}$$

A version of Theorem 19.1.1 applies here too. Each multiplier corresponds to a different derivative of $f\big(\mathbf{x}^*(\mathbf{c})\big)$.

Theorem 19.1.2. *Let $U \subset \mathbb{R}^m$ and suppose $f\colon U \to \mathbb{R}$ and $\mathbf{h}\colon U \to \mathbb{R}^\ell$ are $\mathcal{C}^1$ functions. Let $\mathbf{x}^*(\mathbf{c})$ be the solution to the maximization problem (19.1.4) and $\boldsymbol{\mu}^*(\mathbf{c})$ the corresponding multipliers. Suppose $\mathbf{x}^*(\mathbf{c})$ and $\mu^*(\mathbf{c})$ are $\mathcal{C}^1$ in $\mathbf{c}$ and that NDCQ holds at $\big(\mathbf{x}^*(\mathbf{c}), \boldsymbol{\mu}^*(\mathbf{c})\big)$. Then*

$$\boldsymbol{\mu}^*(\mathbf{c}) = D_{\mathbf{c}} M(\mathbf{c}) = D_{\mathbf{c}}\Big[f\big(\mathbf{x}^*(\mathbf{c})\big)\Big].$$

Proof. The first order conditions are

$$D_{\mathbf{x}} f = \boldsymbol{\mu}^* D_{\mathbf{x}} \mathbf{h} \tag{19.1.5}$$

where everything is evaluated at the optimum. We also differentiate the constraints $\mathbf{h}\big(\mathbf{x}^*(\mathbf{c})\big) = \mathbf{c}$ with respect to $\mathbf{c}$, yielding

$$(D_{\mathbf{x}} \mathbf{h})(D_{\mathbf{c}} \mathbf{x}^*) = \mathbf{I}_{\ell \times \ell} \tag{19.1.6}$$

Use the Chain Rule to differentiate $f\big(\mathbf{x}^*(c), c\big)$ and then substitute equations (19.1.5) and (19.1.6) to complete the proof:

$$\begin{aligned} D_{\mathbf{c}} M(\mathbf{c}) &= D_{\mathbf{c}}\Big[f\big(\mathbf{x}^*(\mathbf{c})\big)\Big] & \\ &= (D_{\mathbf{x}} f)(D_{\mathbf{c}} \mathbf{x}^*) & \text{Chain Rule} \\ &= \boldsymbol{\mu}^*(\mathbf{c})(D_{\mathbf{x}} \mathbf{h})(D_{\mathbf{c}} x^*) & \text{Equation (19.1.5)} \\ &= \boldsymbol{\mu}^*(\mathbf{c}) \mathbf{I}_{\ell \times \ell} & \text{Equation (19.1.6)} \\ &= \boldsymbol{\mu}^*(\mathbf{c}) & \end{aligned}$$

∎

19.1.9 Multipliers and Mixed Constraints

We move on to the general case of mixed constraints. This means we have an optimization problem of this form.

$$\begin{aligned} M(\mathbf{b},\mathbf{c}) = \max_{\mathbf{x}}\ & f(\mathbf{x}) \\ \text{s.t. }\ & g_i(\mathbf{x}) \le b_i, \text{ for } i = 1,\dots,k \\ & h_i(\mathbf{x}) = c_i, \text{ for } i = 1,\dots,\ell \end{aligned} \tag{19.1.7}$$

It's easy to see that if a constraint $g_h(\mathbf{x}^*) \le b_h$ does not bind at $\mathbf{x}^*$, that it will still not bind if we change b_h slightly. It follows that the maximized value, $f\big(\mathbf{x}^*(\mathbf{b},\mathbf{c})\big)$, does not change either, and that its b_h derivative must be zero. Since constraint h does not bind, the corresponding multiplier is zero by complementary slackness. The multiplier and derivative are the same.

19.1.10 Theorem on Multipliers and Mixed Constraints

The proof of the following theorem is similar to the previous results.

BEWARE! The requirement that the maximized value and the maximizers be $\mathcal{C}^1$ can be difficult to meet when there are inequality constraints. The problem is that you might be switching from one type of solution to another when a constraint starts or finishes binding. The transition might lead to a discontinuous derivative.

Theorem 19.1.3. *Let* $U \subset \mathbb{R}^m$ *and suppose* $f\colon U \to \mathbb{R}$, $\mathbf{g}: U \to \mathbb{R}^k$, *and* $\mathbf{h}\colon U \to \mathbb{R}^\ell$ *are* $\mathcal{C}^1$ *functions. Let* $\mathbf{x}^*(\mathbf{b}, \mathbf{c})$ *be the solution to the maximization problem (19.1.7) and* $\boldsymbol{\mu}^*(\mathbf{b}, \mathbf{c})$ *and* $\boldsymbol{\lambda}^*(\mathbf{b}, \mathbf{c})$ *the corresponding multipliers. Suppose* $\mathbf{x}^*(\mathbf{b}, \mathbf{c})$, $\boldsymbol{\mu}^*(\mathbf{b}, \mathbf{c})$ *and* $\boldsymbol{\lambda}^*(\mathbf{b}, \mathbf{c})$ *are* $\mathcal{C}^1$ *in* $(\mathbf{b}, \mathbf{c})$.

Let $\hat{\mathbf{g}}$ *be the vector of the* $\hat{k}$ *binding constraints at* $\mathbf{x}^*$ *and define*

$$\mathbf{G}(\mathbf{x}) = \begin{pmatrix} \hat{\mathbf{g}}(\mathbf{x}) \\ \mathbf{h}(\mathbf{x}) \end{pmatrix}.$$

Suppose $\operatorname{rank} D\mathbf{G}(\mathbf{x}^*) = \hat{k} + \ell$ *holds (NDCQ).*

Then

$$\boldsymbol{\lambda}^*(\mathbf{b}, \mathbf{c}) = D_{\mathbf{b}} M(\mathbf{b}, \mathbf{c}) = D_{\mathbf{b}}\Big[f\big(\mathbf{x}^*(\mathbf{b}, \mathbf{c})\big)\Big]$$

$$\boldsymbol{\mu}^*(\mathbf{b}, \mathbf{c}) = D_{\mathbf{c}} M(\mathbf{b}, \mathbf{c}) = D_{\mathbf{c}}\Big[f\big(\mathbf{x}^*(\mathbf{b}, \mathbf{c})\big)\Big].$$

19.1.11 Envelope Theorem

In the above theorem, we've been able to calculate the marginal value of relaxing any of the constraints. However, we would like to calculate marginal values for other parameters of the model, such as prices. For that, we need a generalization known as the Envelope Theorem.

Let $\mathbf{a} \in \mathbb{R}^n$ be a vector of parameters. Suppose the objective depends on both $\mathbf{x}$ and $\mathbf{a}$, with $f(\mathbf{x}, \mathbf{a}) \in \mathcal{C}^1$ and suppose we can also write $\mathbf{h}(\mathbf{x}, \mathbf{a}) \in \mathcal{C}^1$ where $\mathbf{h}(\mathbf{x}, \mathbf{a}) = \mathbf{0}$ is the vector of equality constraints.

Let $M(\mathbf{a}) = f(\mathbf{x}^*(\mathbf{a}), \mathbf{a})$ be the maximized value of f.

19.1.12 Redoing the Standard Calculations

The Lagrangian is

$$\mathcal{L} = f(\mathbf{x}, \mathbf{a}) - \boldsymbol{\mu}\mathbf{h}(\mathbf{x}, \mathbf{a})$$

We now employ the same procedure as before. The first order conditions are

$$0 = D_{\mathbf{x}}\mathcal{L} = D_{\mathbf{x}}f - \boldsymbol{\mu}^* D_{\mathbf{x}}\mathbf{h},$$

so

$$D_{\mathbf{x}}f = \boldsymbol{\mu}^* D_{\mathbf{x}}\mathbf{h}. \tag{19.1.8}$$

We differentiate the constraint, which now has an extra term

$$D_{\mathbf{a}}\Big[\mathbf{h}\big(\mathbf{x}^*(\mathbf{a}), \mathbf{a}\big)\Big] = (D_{\mathbf{x}}\mathbf{h})\,(D_{\mathbf{a}}\mathbf{x}^*) + D_{\mathbf{a}}\mathbf{h} = \mathbf{0}$$

so

$$(D_{\mathbf{x}}\mathbf{h})\,(D_{\mathbf{a}}\mathbf{x}^*) = -D_{\mathbf{a}}\mathbf{h} \tag{19.1.9}$$

Then we wrap it all together using the Chain Rule and substitute equations (19.1.8) and (19.1.9)

$$\begin{aligned}
D_{\mathbf{a}}M(\mathbf{a}) &= D_{\mathbf{a}}\Big[f\big(\mathbf{x}^*(\mathbf{a}), \mathbf{a}\big)\Big] && \text{Definition of } M\\
&= (D_{\mathbf{x}}f)\,(D_{\mathbf{a}}\mathbf{x}^*) + D_{\mathbf{a}}f && \text{Chain Rule}\\
&= \big(\boldsymbol{\mu}^* D_{\mathbf{x}}\mathbf{h}\big)D_{\mathbf{a}}\mathbf{x}^* + D_{\mathbf{a}}f && \text{Equation (19.1.8)}\\
&= -\boldsymbol{\mu}^* D_{\mathbf{a}}\mathbf{h} + D_{\mathbf{a}}f && \text{Equation (19.1.9)}\\
&= D_{\mathbf{a}}\big[f - \boldsymbol{\mu}^*\mathbf{h}\big] && \text{Rearrangement}\\
&= D_{\mathbf{a}}\mathcal{L}\big(\mathbf{x}^*(\mathbf{a}), \boldsymbol{\mu}^*(\mathbf{a}), \mathbf{a}\big) && \text{Definition of } \mathcal{L}
\end{aligned}$$

19.1.13 Statement of Envelope Theorem

We now state an Envelope Theorem based on the above calculations. We just need to add appropriate hypotheses so that the preceding calculations are valid. In particular, the hypotheses must guarantee that our first order conditions are correct and that the maximizers are $\mathcal{C}^1$.

Envelope Theorem. *Consider the maximization problem*

$$\begin{aligned} M(\mathbf{a}) = \max_{\mathbf{x}}\ & f(\mathbf{x}, \mathbf{a}) \\ \textit{s.t.}\ & h_j(\mathbf{x}, \mathbf{a}) = 0, \ \textit{for}\ j = 1, \dots, \ell \end{aligned} \tag{19.1.10}$$

Let $U \subset \mathbb{R}^m$ *and* $V \subset \mathbb{R}^n$ *and suppose* $f\colon U \times V \to \mathbb{R}$ *and* $\mathbf{h}\colon U \times V \to \mathbb{R}^\ell$ *are* $\mathcal{C}^1$ *functions. Let* $\mathbf{x}^*(\mathbf{a})$ *solve the maximization problem (19.1.10) for any* $\mathbf{a} \in V$*. Define the Lagrangian*

$$\mathcal{L} = f(\mathbf{x}, \mathbf{a}) - \boldsymbol{\mu}\mathbf{h}(\mathbf{x}, \mathbf{a}).$$

Suppose the maximizer $\mathbf{x}^*(\mathbf{a})$ *and the corresponding Lagrange multipliers* $\boldsymbol{\mu}^*(\mathbf{a})$ *are* $\mathcal{C}^1$ *functions of* $\mathbf{a}$*, and that NDCQ holds. Then*

$$D_{\mathbf{a}} M(\mathbf{a}) = D_{\mathbf{a}} \mathcal{L}\big(\mathbf{x}^*(\mathbf{a}), \boldsymbol{\mu}^*(\mathbf{a}), \mathbf{a}\big).$$

Theorem 19.1.2 is just the special case with $\mathbf{h}(\mathbf{x}, \mathbf{a}) = \mathbf{h}(\mathbf{x}) - \mathbf{a}$.

19.1.14 Indirect Utility and the Envelope Theorem

Consider a simple consumer's problem.

▶ Example 19.1.4: Roy's Identity. The indirect utility function $v(\mathbf{p}, m)$ is defined by

$$v(\mathbf{p}, m) = \max_{\mathbf{x}} u(\mathbf{x})$$
$$\text{s.t. } \mathbf{p}\cdot\mathbf{x} = m$$

where $u\colon \mathbb{R}^m_+ \to \mathbb{R}$ is a utility function, the prices obey $\mathbf{p} \gg \mathbf{0}$ and income is positive, $m > 0$.

We set $\mathbf{a} = \mathbf{p}$. Then

$$\frac{\partial v}{\partial p_i} = \big(D_{\mathbf{p}}\mathcal{L}\big)_i = \frac{\partial u}{\partial p_i} - \mu^*(\mathbf{a})\, x_i(\mathbf{p}, m) = -\mu^*(\mathbf{a})\, x_i(\mathbf{p}, m).$$

We can use the fact that $\partial v/\partial m = \mu^*$ to establish *Roy's Identity*:[2]

$$x_i(\mathbf{p}, m) = -\frac{\partial v/\partial p_i}{\partial v/\partial m}.$$

◀

[2] Roy's identity was established by the French economist and econometrican René Roy (1894–1977). See René Roy (1947), La distribution du revenu entre les divers biens, *Econometrica* **15**, 205–225.

19.1.15 Cost and the Envelope Theorem

Another example involves the cost function.[3]

▶ **Example 19.1.5: Shephard's Lemma.** The cost function $c(\boldsymbol{w}, q)$ is defined by

$$\begin{aligned} c(\boldsymbol{w}, q) = \min_{\boldsymbol{z}} \ & \boldsymbol{w} \cdot \boldsymbol{z} \\ \text{s.t. } & f(\boldsymbol{z}) = q, . \end{aligned}$$

where $f\colon \mathbb{R}^m_+ \to \mathbb{R}$ is a production function and $\boldsymbol{w} \gg \mathbf{0}$ is the vector of factor prices.

Since $\boldsymbol{w}$ doesn't appear in the constraints, the Envelope Theorem tells us that the derivative of the cost function is

$$D_{\boldsymbol{w}} c(\boldsymbol{w}, q) = D_{\boldsymbol{w}} \mathcal{L} = D_{\boldsymbol{w}}(\boldsymbol{w} \cdot \boldsymbol{z}) = \boldsymbol{z}(\boldsymbol{w}, q)^{\mathsf{T}}.$$

The row vector $\boldsymbol{z}(\boldsymbol{w}, q)^{\mathsf{T}}$ is the vector of *conditional factor demands*. This result is known as *Shephard's Lemma*. ◀

[3] Shephard's Lemma is due to Ronald W. Shephard (1953), *Cost and Production Functions*, Princeton University Press, Princeton, NJ. Cost functions had been used prior to Shephard, but he may have been the first to use the modern definition. He used a duality method to prove Shephard's Lemma, based on the distance function.

Ronald W. Shephard (1912–1982) was an American economist who's research concentrated on cost and production. He pioneered the use of duality based on the distance function. In contrast, the usual economic duality is more directly related to the conjugate function.

19.1.16 Expenditure and the Envelope Theorem

We can also apply this to the very similar expenditure function.[4]

▶ **Example 19.1.6: Shephard-McKenzie Lemma.** The expenditure function solves

$$e(\mathbf{p}, \bar{u}) = \min_{\mathbf{x}} \mathbf{p}\cdot\mathbf{x}$$
$$\text{s.t. } u(\mathbf{x}) = \bar{u}, .$$

By the Envelope Theorem,

$$D_{\mathbf{p}} e(\mathbf{p}, \bar{u}) = D_{\mathbf{p}} \mathcal{L} = D_{\mathbf{p}}(\mathbf{p}\cdot\mathbf{x}) = \mathbf{h}(\mathbf{p}, \bar{u})^{\mathsf{T}},$$

yielding the *Hicksian* or *compensated demands* $\mathbf{h}(\mathbf{p}, \bar{u})$. This result is known as the *Shephard-McKenzie Lemma*. ◀

[4] The expenditure function was introduced by Lionel McKenzie (1957), Demand theory without a utility index, *Rev. Econ. Studies* **24**, 185–189.

Lionel McKenzie (1919–2010) was an American economist known for proving the existence of general equilibrium, for his work on demand theory, and for optimal growth and turnpike theory.

19.2 Second Order Necessary and Sufficient Conditions

The problem with merely looking at the first order conditions is that they are not enough to ensure that we have a maximum or minimum. They are necessary, but not sufficient.

As we will see, there are second order conditions for constrained problems that are sufficient, but not necessary. As far as necessity is concerned, the best we can do is find conditions that are locally necessary.

19.2.1 Constrained Negative Definite Lagrangians

As in the unconstrained case, we can use the second derivative to show that a critical point of (19.1.4) is maximal. For the sufficient conditions, we need to show that $D^2_{\mathbf{x}}\mathcal{L}$ is negative definite, at least for those vectors that are in the tangent space of the constraint set.

We can apply the bordered Hessian test to see if $D^2_{\mathbf{x}}\mathcal{L}$ is negative definite on the tangent space, $T_{\mathbf{x}^*}M = \ker D_{\mathbf{x}^*}\mathbf{h}$. Define the bordered Hessian of the Lagrangian by

$$\mathbf{B} = \begin{pmatrix} 0_{\ell\times\ell} & D\mathbf{h}(\mathbf{x}^*) \\ D\mathbf{h}(\mathbf{x}^*)^T & D^2_{\mathbf{x}}\mathcal{L} \end{pmatrix}.$$

where $0_{\ell\times\ell}$ is the $\ell \times \ell$ zero matrix.

19.2.2 Second Order Sufficient Conditions: Maximum

We can now state second order sufficient conditions for a constrained strict local maximum.

Theorem 19.2.1. *Let* $U \subset \mathbb{R}^m$ *and suppose* $f\colon U \to \mathbb{R}$ *and* $\mathbf{h}\colon U \to \mathbb{R}^\ell$ *are* $\mathcal{C}^2$ *functions. Form the Lagrangian*

$$\mathcal{L} = f(\mathbf{x}) - \boldsymbol{\mu}\big(\mathbf{h}(\mathbf{x}) - \mathbf{c}\big).$$

Suppose that the following conditions are satisfied:

1. $\mathbf{x}^* \in M = \{\mathbf{x} : \mathbf{h}(\mathbf{x}) = \mathbf{c}\}$.
2. $D_{\mathbf{x}}\mathcal{L} = \mathbf{0}$ *and* $D_{\boldsymbol{\mu}}\mathcal{L} = \mathbf{0}$ *at* $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ *(critical point).*
3. $\operatorname{rank} D\mathbf{h}(\mathbf{x}^*) = \ell$ *holds (NDCQ).*
4. *The Hessian of the Lagrangian with respect to* $\mathbf{x}$ *at* $(\mathbf{x}^*, \boldsymbol{\mu}^*)$, $D^2_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*)$, *is negative definite on the tangent space* $T_{\mathbf{x}^*}M = \ker D_{\mathbf{x}^*}\mathbf{h} = \{\mathbf{v} : \big(D_{\mathbf{x}^*}\mathbf{h}\big)\mathbf{v} = \mathbf{0}\}$.

Assumptions (1) and (2) tell us that $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ *is a critical point of the problem (19.1.4).*

Then $\mathbf{x}^*$ *is a strict local maximum of* f *on the constraint set* M.

Moreover, the conclusion remains true if (4) is replaced by (4′).

4′. *The last* $(m - \ell)$ *leading principal minors of the bordered Hessian* $\mathbf{B}$ *alternate in sign and the determinant of* $\mathbf{B}$ *obeys* $(-1)^{\ell+m} \det \mathbf{B} > 0$.

19.2.3 The Natural Bordered Hessian

It is arguably more natural to use $D^2_{(\mu,x)}\mathcal{L}$ rather than the bordered Hessian. Here

$$D^2_{(\mu,x)}\mathcal{L}(\mathbf{x}^*,\boldsymbol{\mu}^*) = \begin{pmatrix} \mathbf{0}_{\ell\times\ell} & -D\mathbf{h}(\mathbf{x}^*) \\ -D\mathbf{h}(\mathbf{x}^*)^T & D^2_x\mathcal{L}(\mathbf{x}^*,\boldsymbol{\mu}^*) \end{pmatrix}$$

The extra minus signs do not affect whether the matrix $D^2_x\mathcal{L}$ is constrained positive definite or constrained negative definite. Compared to the usual bordered Hessian, the principal minors are unchanged because for each column that is multiplied by (-1), the corresponding row is also multiplied by (-1). Together, they multiply each principal minor by $(-1)^2$, leaving every principal minor unchanged.

The bordered Hessian that appears in the proof of Theorem 19.2.1 is the natural bordered Hessian, $D^2_{(\mu,x)}\mathcal{L}$, not our usual bordered Hessian. Since they are equivalent in the sense that they both are negative definite on the same subspaces and yield the same conditions on minors, it doesn't really affect the statement of the theorem.

19.2.4 Proof of Theorem 19.2.1

Proof. We start by noting that (4) and (4′) are equivalent. It doesn't matter which we use.

The rest of the proof proceeds by contradiction. **Suppose that x^* is not a strict local maximum**. Then for $n = 1, 2, \ldots,$ we can find an $\mathbf{x}_n \neq \mathbf{x}^*$ obeying $\mathbf{h}(\mathbf{x}_n) = \mathbf{c}$ with $\|\mathbf{x}_n - \mathbf{x}^*\| < 1/n$ and $f(\mathbf{x}_n) \geq f(\mathbf{x}^*)$.

Next define a sequence of unit vectors by

$$\mathbf{u}_j = \frac{\mathbf{x}_j - \mathbf{x}^*}{\|\mathbf{x}_j - \mathbf{x}^*\|}.$$

These all lie on the unit sphere $S^{m-1} = \{\mathbf{x} \in \mathbb{R}^m : \|\mathbf{x}\| = 1\}$, which is a compact set. It follows that there is a convergent subsequence $\mathbf{u}_{j_k}$. We call its limit $\mathbf{u}^*$.

Since $\mathbf{h}$ is $\mathcal{C}^2$, its first order Taylor polynomial with remainder about $\mathbf{x}^*$ can be written[5]

$$\mathbf{h}(\mathbf{x}_{j_k}) = \mathbf{h}(\mathbf{x}^*) + D\mathbf{h}(\mathbf{x}^*)(\mathbf{x}_{j_k} - \mathbf{x}^*) + R_1(\mathbf{x}_{k_j}; \mathbf{x}^*) \tag{19.2.11}$$

where $R_1(\mathbf{x}_{k_j}; \mathbf{x}^*) = o\big(\|\mathbf{x}_{k_j} - \mathbf{x}^*\|\big)$.

Using the fact $\mathbf{h}(\mathbf{x}_{j_k}) = \mathbf{c} = \mathbf{h}(\mathbf{x}^*)$, we can simplify equation (19.2.11), and then divide by $\|\mathbf{x}_{j_k} - \mathbf{x}^*\|$, obtaining

$$0 = D\mathbf{h}(\mathbf{x}^*)\left(\frac{\mathbf{x}_{j_k} - \mathbf{x}^*}{\|\mathbf{x}_{j_k} - \mathbf{x}^*\|}\right) + \frac{R_1(\mathbf{x}_{k_j}; \mathbf{x}^*)}{\|\mathbf{x}_{j_k} - \mathbf{x}^*\|}.$$

Taking the limit as $k \to \infty$ yields

$$0 = D\mathbf{h}(\mathbf{x}^*)\mathbf{u}^*$$

showing that $\mathbf{u}^* \in T_{\mathbf{x}^*}M$.

Proof continues ...

[5] See section 30.29.

19.2.5 Remainder of Proof of Theorem 19.2.1

Remainder of Proof. Take the second order Taylor polynomial of the Lagrangian $\mathcal{L}$ about $\mathbf{x}^*$ with remainder S_2.

$$\begin{aligned}\mathcal{L}(\mathbf{x}_{j_k}, \boldsymbol{\mu}^*) &= \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*) + D_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*)(\mathbf{x}_{j_k} - \mathbf{x}^*) \\ &\quad + \frac{1}{2}(\mathbf{x}_{j_k} - \mathbf{x}^*)^{\mathsf{T}}\left[D^2_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*)\right](\mathbf{x}_{j_k} - \mathbf{x}^*) + S_2(\mathbf{x}_{j_k}; \mathbf{x}^*)\end{aligned}$$

where $S_2(\mathbf{x}_{j_k} - \mathbf{x}^*) = o\big(\|\mathbf{x}_{j_k} - \mathbf{x}^*\|^2\big)$.

By the first order condition, $D_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*) = \mathbf{0}$, so

$$\mathcal{L}(\mathbf{x}_{j_k}, \boldsymbol{\mu}^*) - \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*) = \frac{1}{2}(\mathbf{x}_{j_k} - \mathbf{x}^*)^{\mathsf{T}}\left[D^2_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*)\right](\mathbf{x}_{j_k} - \mathbf{x}^*) + S_2(\mathbf{x}_{j_k}; \mathbf{x}^*).$$

Using the facts that $\mathbf{h}(\mathbf{x}_{j_k}) = \mathbf{c} = \mathbf{h}(\mathbf{x}^*)$ for all $k = 1, 2, 3 \ldots,$ we find $\mathcal{L}(\mathbf{x}_{j_k}, \boldsymbol{\mu}^*) = f(\mathbf{x}_{j_k})$ and $\mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*) = f(\mathbf{x}^*)$. **By the contradiction hypothesis**,

$$\begin{aligned}0 &\le f(\mathbf{x}_{j_k}) - f(\mathbf{x}^*) \\ &\le \mathcal{L}(\mathbf{x}_{j_k}, \boldsymbol{\mu}^*) - \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*) \\ &= \frac{1}{2}(\mathbf{x}_{j_k} - \mathbf{x}^*)^{\mathsf{T}}\left[D^2_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*)\right](\mathbf{x}_{j_k} - \mathbf{x}^*) + S_2(\mathbf{x}_{j_k}; \mathbf{x}^*).\end{aligned}$$

Dividing by $\|\mathbf{x}_{j_k} - \mathbf{x}^*\|^2$ and letting $k \to \infty$, we find

$$0 \le \frac{1}{2}\mathbf{u}^{*\mathsf{T}}\left[D^2_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*)\right]\mathbf{u}^*.$$

But $\mathbf{u}^*$ is a unit vector with $\mathbf{u}^* \in T_{\mathbf{x}^*}M$, and $D^2_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*)$ is negative definite on such vectors by hypothesis. The right-hand side of the equation is negative, so $0 < 0$, which is impossible. **This contradiction shows** that $\mathbf{x}^*$ is a strict local max.

19.2.6 Second Order Sufficient Conditions: Minimum

Let $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ be a critical point of the minimization problem analogous to (19.1.4). If the Hessian is positive definite on $T_{\mathbf{x}^*}M = \ker D_{\mathbf{x}^*}\mathbf{h}$ at a critical point $\mathbf{x}^*$, then $\mathbf{x}^*$ is a minimum.

Theorem 19.2.2. *Let $U \subset \mathbb{R}^m$ and suppose $f\colon U \to \mathbb{R}$ and $\mathbf{h}\colon U \to \mathbb{R}^\ell$ are $\mathcal{C}^2$ functions. Form the Lagrangian*

$$\mathcal{L} = f(\mathbf{x}) - \boldsymbol{\mu}\big(\mathbf{c} - \mathbf{h}(\mathbf{x})\big).$$

Suppose that the following conditions are satisfied:

1. $\mathbf{x}^* \in M = \{\mathbf{x} : \mathbf{h}(\mathbf{x}) = \mathbf{c}\}$.
2. $D_{\mathbf{x}}\mathcal{L} = \mathbf{0}$ *and* $D_{\boldsymbol{\mu}}\mathcal{L} = \mathbf{0}$ *at* $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ *(critical point).*
3. $\operatorname{rank} D\mathbf{h}(\mathbf{x}^*) = \ell$ *holds (NDCQ).*
4. *The Hessian of the Lagrangian with respect to* $\mathbf{x}$ *at* $(\mathbf{x}^*, \boldsymbol{\mu}^*)$, $D^2_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*)$ *is positive definite on the tangent space* $T_{\mathbf{x}^*}M = \ker D_{\mathbf{x}^*}\mathbf{h} = \{\mathbf{v} : \big(D_{\mathbf{x}^*}\mathbf{h}\big)\mathbf{v} = \mathbf{0}\}$.

Assumptions (1) and (2) tell us that $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ is a critical point of the minimization problem analogous to (19.1.4).

Then $\mathbf{x}^*$ is a strict local constrained minimum of f on M.

Moreover, the conclusion remains true if (4) is replaced by (4′).

4′. *The last $(m - \ell)$ leading principal minors of the bordered Hessian $\mathbf{B}$ have the same sign and the determinant of $\mathbf{B}$ obeys $(-1)^\ell \det \mathbf{B} > 0$.*

Proof. Mimic the proof of Theorem 19.2.1. ∎

19.2.7 2nd Order Conditions: Mixed Constraint Maxima

We'll jump directly to the full mixed constraint version of the theorem.

Theorem 19.2.3. *Let $U \subset \mathbb{R}^m$ and suppose the functions $f\colon U \to \mathbb{R}$, $\mathbf{g}: U \to \mathbb{R}^k$, and $\mathbf{h}\colon U \to \mathbb{R}^\ell$ are $\mathcal{C}^2$ functions. Suppose further that there are $\hat{k}$ binding constraints at $\mathbf{x}^*$. Let $\hat{\mathbf{g}}$ be the vector of binding inequality constraints at $\mathbf{x}^*$ with corresponding constants $\hat{\mathbf{b}}$. Set*

$$\mathbf{G}(\mathbf{x}) = \begin{pmatrix} \hat{\mathbf{g}}(\mathbf{x}) \\ \mathbf{h}(\mathbf{x}) \end{pmatrix}$$

Suppose $\operatorname{rank} D\mathbf{G}(\mathbf{x}^*) = \hat{k} + \ell$ *holds (NDCQ) and define the constraint set by* $M = \{\mathbf{x} : \mathbf{g}(\mathbf{x}) \le \mathbf{b}, \mathbf{h}(\mathbf{x}) = \mathbf{c}\}$. *Form the Lagrangian:*

$$\mathcal{L} = f(\mathbf{x}) - \boldsymbol{\lambda}\big(\hat{\mathbf{g}}(\mathbf{x}) - \hat{\mathbf{b}}\big) - \boldsymbol{\mu}\big(\mathbf{h}(\mathbf{x}) - \mathbf{c}\big).$$

Now suppose

1. $\mathbf{x}^* \in M$.
2. *There are $\boldsymbol{\lambda}^* \in \mathbb{R}^k$ and $\boldsymbol{\mu}^* \in \mathbb{R}^\ell$ with*

$$\begin{gathered} D_\mathbf{x}\mathcal{L} = \mathbf{0} \quad \textit{at } (\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*), \\ \boldsymbol{\lambda}^* \ge \mathbf{0}, \\ \lambda_1^*\big(g_1(\mathbf{x}^*) - b_1\big) = 0, \dots, \lambda_k^*\big(g_k(\mathbf{x}^*) - b_k\big) = 0, \\ \mathbf{h}(\mathbf{x}^*) = \mathbf{c}. \end{gathered}$$

3. *The Hessian of the Lagrangian with respect to $\mathbf{x}$ at $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, $D^2_\mathbf{x}\mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is negative definite on the tangent space $T_{\mathbf{x}^*}M = \ker D_{\mathbf{x}^*}\mathbf{G} = \{\mathbf{v} : \big(D_{\mathbf{x}^*}\mathbf{G}\big)\mathbf{v} = \mathbf{0}\}$.*

Assumptions (1) and (2) tell us $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ is a critical point of the maximization problem (19.1.4).

Then $\mathbf{x}^*$ is a strict local constrained maximum of f on M.

Moreover, the conclusion remains true if (3) is replaced by (3′).

3′. *The last $(m - \hat{k} - \ell)$ leading principal minors of the bordered Hessian $\mathbf{B}$ alternate in sign and the determinant of $\mathbf{B}$ obeys $(-1)^m \det \mathbf{B} > 0$.*

19.2.8 2nd Order Conditions: Mixed Constraint Minima

The version for minima is only slightly different. Although the differences are minor, they make all the difference in the world, resulting in minima rather than maxima.

Theorem 19.2.4. *Let $U \subset \mathbb{R}^m$ and suppose the functions $f\colon U \to \mathbb{R}$, $\mathbf{g}: U \to \mathbb{R}^k$, and $\mathbf{h}\colon U \to \mathbb{R}^\ell$ are $\mathcal{C}^2$ functions. Suppose further that there are $\hat{k}$ binding constraints at $\mathbf{x}^*$. Let $\hat{\mathbf{g}}$ be the vector of binding inequality constraints at $\mathbf{x}^*$ with corresponding constants $\hat{\mathbf{b}}$. Set*

$$\mathbf{G}(\mathbf{x}) = \begin{pmatrix} \hat{\mathbf{g}}(\mathbf{x}) \\ \mathbf{h}(\mathbf{x}) \end{pmatrix}$$

Suppose $\operatorname{rank} D\mathbf{G}(\mathbf{x}^*) = \hat{k} + \ell$ *holds (NDCQ) and define the constraint set by $M = \{\mathbf{x} : \mathbf{g}(\mathbf{x}) \ge \mathbf{b}, \mathbf{h}(\mathbf{x}) = \mathbf{c}\}$. Form the Lagrangian:*

$$\mathcal{L} = f(\mathbf{x}) - \boldsymbol{\lambda}\big(\hat{\mathbf{g}}(\mathbf{x}) - \mathbf{b}\big) - \boldsymbol{\mu}\big(\mathbf{h}(\mathbf{x}) - \mathbf{c}\big).$$

Now suppose

1. $\mathbf{x}^* \in M$.
2. *There are $\boldsymbol{\lambda}^* \in \mathbb{R}^k$ and $\boldsymbol{\mu}^* \in \mathbb{R}^\ell$ with*

$$\begin{gathered} D_{\mathbf{x}}\mathcal{L} = \mathbf{0} \quad \textit{at } (\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*), \\ \boldsymbol{\lambda}^* \ge \mathbf{0}, \\ \lambda_1^*\big(g_1(\mathbf{x}^*) - b_1\big) = 0, \dots, \lambda_k^*\big(g_k(\mathbf{x}^*) - b_k\big) = 0, \\ \mathbf{h}(\mathbf{x}^*) = \mathbf{c}. \end{gathered}$$

3. *The Hessian of the Lagrangian with respect to $\mathbf{x}$ at $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, $D_{\mathbf{x}}^2\mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is positive definite on the tangent space $T_{\mathbf{x}^*}M =$* $\ker D_{\mathbf{x}^*}\mathbf{G} = \{\boldsymbol{v} : \big(D_{\mathbf{x}^*}\mathbf{G}\big)\boldsymbol{v} = \mathbf{0}\}$.

Then $\mathbf{x}^*$ is a strict local constrained minimum of f on M.

Moreover, the conclusion remains true if (3) is replaced by (3′).

3′. *The last $(m - \hat{k} - \ell)$ leading principal minors of the bordered Hessian $\mathbf{B}$ have the same sign and the determinant of $\mathbf{B}$ obeys* $(-1)^{\hat{k}+\ell} \det \mathbf{B} > 0$.

19.2.9 Second Order Necessary Conditions, Preliminaries

We will limit ourselves to a single theorem giving necessary conditions for a local maximum in the mixed case. The local minimum case is similar. This involves semidefinite matrices. There are also sufficient conditions using semidefinite matrices, analogous to Theorem 17.8.1 and Theorem 17.8.1 Restated.

Let $\hat{k}$ be the number of binding inequality constraints at $\mathbf{x}^*$ and $\hat{\mathbf{g}}$ the corresponding vector of binding inequality constraints. Form the $\hat{k} + \ell$-vector of binding constraints at $\mathbf{x}^*$,

$$\mathbf{G}(\mathbf{x}) = \begin{pmatrix} \hat{\mathbf{g}}(\mathbf{x}) \\ \mathbf{h}(\mathbf{x}) \end{pmatrix}$$

Consider the bordered Hessian of the Lagrangian:

$$\mathbf{B} = \begin{pmatrix} \mathbf{0}_{(\hat{k}+\ell)\times(\hat{k}+\ell)} & D\mathbf{G}(\mathbf{x}^*) \\ D\mathbf{G}(\mathbf{x}^*)^T & D^2_{\mathbf{x}^*}\mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \end{pmatrix}$$

Bordered Principal Submatrices. Let $\hat{\mathbf{B}}_r$ denote any $r \times r$ principal submatrix with $r > \hat{k} + \ell$, containing the $\mathbf{0}_{(\hat{k}+\ell)\times(\hat{k}+\ell)}$ block in upper left hand corner of $\mathbf{B}$. We will refer to these as *bordered principal submatrices of rank* r.

19.2.10 Second Order Necessary Conditions

We can now state the theorem.

Theorem 19.2.5. *Let* $U \subset \mathbb{R}^m$ *and suppose the functions* $f\colon U \to \mathbb{R}$, $\mathbf{g}\colon U \to \mathbb{R}^k$, *and* $\mathbf{h}\colon U \to \mathbb{R}^\ell$ *are* $\mathcal{C}^2$ *functions. Suppose further that there are* $\hat{k}$ *binding inequality constraints at* $\mathbf{x}^*$. *Let* $\hat{\mathbf{g}}$ *be the vector of binding inequality constraint functions at* $\mathbf{x}^*$ *with corresponding constants* $\hat{\mathbf{b}}$. *Then set*

$$\mathbf{G}(\mathbf{x}) = \begin{pmatrix} \hat{\mathbf{g}}(\mathbf{x}) \\ \mathbf{h}(\mathbf{x}) \end{pmatrix},$$

which includes all binding constraints at $\mathbf{x}^*$.

Suppose that $\operatorname{rank} D\mathbf{G}(\mathbf{x}^*) = \hat{k} + \ell$ *holds (NDCQ) and define the constraint set by* $M = \{\mathbf{x} : \mathbf{g}(\mathbf{x}) \le \mathbf{b}, \mathbf{h}(\mathbf{x}) = \mathbf{c}\}$. *Suppose that* $\mathbf{x}^* \in U^0$ *is a local maximum of* f *on* M, *that* $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ *is a critical point for the Lagrangian*

$$\mathcal{L} = f(\mathbf{x}) - \boldsymbol{\lambda}\big(\hat{\mathbf{g}}(\mathbf{x}) - \hat{\mathbf{b}}\big) - \boldsymbol{\mu}\big(\mathbf{h}(\mathbf{x}) - \mathbf{c}\big),$$

and that the complementary slackness and non-negativity conditions hold.

Then the Hessian of the Lagrangian at $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ *is negative semidefinite on the tangent space* $T_{\mathbf{x}^*} = \ker D\mathbf{G}(\mathbf{x}^*) = \{\mathbf{v} \in \mathbb{R}^{h+\ell} : \big(D\mathbf{G}(\mathbf{x}^*)\big)\mathbf{v} = \mathbf{0}\}$.

Equivalently, either the minors $\det \hat{\mathbf{B}}_r$ *of any given size* $r > m-(\hat{k}+\ell)$ *are either all non-positive, or all non-negative. The generalized sign alternates with* r, *and* $(-1)^{\hat{k}+\ell} \det \hat{\mathbf{B}} > 0$.

Simon and Blume punted on the determinant conditions. The best thing to do is to look at the source. Our determinant condition is a rewritten version of the one in Debreu (1952).[6]

Minimum. If $\mathbf{x}^*$ is a minimum, we have to reverse the inequalities for $\mathbf{g}$, and adjust the Lagrangian accordingly. In that case, we can conclude that the Hessian of the Lagrangian at $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is positive semidefinite on the tangent space $T_{\mathbf{x}^*}$.

Equivalently, the minors $\det \hat{\mathbf{B}}_r$ with $r > m - (\hat{k} + \ell)$ are either all non-positive, or all non-negative. and $(-1)^{m-\hat{k}-\ell} \det \hat{\mathbf{B}} > 0$.

[6] Debreu, Gerard (1952) Definite and semidefinite quadratic forms, *Econometrica*, **20**, 295–300.

19.3 Smoothness and Optimization

Earlier, when considering the Envelope Theorem and related results, we required that the maximum or minimum be a $\mathcal{C}^1$ function of various parameters. One weakness of that approach is that we gave no method of ensuring that the functions were $\mathcal{C}^1$.

We are ready to remedy that and will give conditions on the primitives of the model, the objectives and constraints, that guarantee the solutions are $\mathcal{C}^1$. The Implicit Function Theorem is the key tool for this.

19.3.1 Parameters and Optima

We start with an unconstrained problem to explore the basic ideas. Consider a simple model with a single parameter a, and two variables, $(x, y) \in U \subset \mathbb{R}^2$. We consider the maximization problem

$$F(a) = \max_{(x,y)\in U} f(x, y, a).$$

If U is compact, the Weierstrass Theorem guarantees a solution. If we have an interior maximum, we know that the first order conditions must be satisfied,

$$\frac{\partial f}{\partial x} = 0 \quad \text{and} \quad \frac{\partial f}{\partial y} = 0.$$

These equations may implicitly define functions $x^*(a)$ and $y^*(a)$ that maximize f for at any given a.

19.3.2 What is Required for Smoothness?

For this to work, the Implicit Function Theorem requires that the (x, y)-derivative of this vector equation be non-singular at $\big(x^*(a), y^*(a), a\big)$. That is, the Hessian matrix

$$D^2_{(x,y)} f\big(\mathbf{x}(a), y(a), a\big)$$

must be non-singular.

If the second order sufficient conditions for a strict local maximum are satisfied, the Hessian $D^2_{(x,y)} f$ will be negative definite, and hence non-singular.

So let's write a theorem summing up what we proved. The Implicit Function Theorem requires that our implicit function be defined by a function that is itself $\mathcal{C}^1$. It also allows multiple exogenous variables, like our variable a.

Differentiability is a local property, so the fact that the Implicit Function Theorem only gives local results should not be a problem.

If we consider a similar minimization problem, the only real difference is that the similar second order sufficient condition is that $D^2_{\mathbf{x}} f$ be positive definite at $\mathbf{x}(a)$, which again means $D^2_{\mathbf{x}} f$ is non-singular.

Let's add these considerations to our theorem.

19.3.3 Smoothness of Unconstrained Optima

Here's a version of such a theorem, stated for $\mathbf{x} \in \mathbb{R}^m$.

Theorem 19.3.1. *Let* $U \subset \mathbb{R}^m$ *and* $V \subset \mathbb{R}$*. Suppose* $f\colon U \times V \to \mathbb{R}$ *is* $\mathcal{C}^2$ *in* $\mathbf{x} \in U$ *and* $\mathcal{C}^1$ *in* $\mathbf{a} \in V$*. Suppose also there are* $(\mathbf{x}_0, a_0) \in U^0 \times V^0$ *with* $D_{\mathbf{x}} f(\mathbf{x}_0, \mathbf{a}_0) = 0$*.*

If the Hessian $D^2_{\mathbf{x}} f(\mathbf{x}_0, a_0)$ *is either positive definite or negative definite, then there is an* $\varepsilon > 0$ *and a* $\mathcal{C}^1$ *function* $\mathbf{x}^*\colon B_\varepsilon(a_0) \to \mathbb{R}^m$ *with* $\mathbf{x}^*(a)$ *obeying*

$$D_{\mathbf{x}} f\big(\mathbf{x}(\mathbf{a}), \mathbf{a}\big) = 0$$

on $B_\varepsilon(\mathbf{a}_0)$ *with* $\mathbf{x}(\mathbf{a}_0) = \mathbf{x}_0$*.*

The function maximizes f *if the Hessian is negative definite and minimizes* f *if the Hessian is positive definite.*

Proof. Left as an exercise to the reader. All of the key components were mentioned on the previous page or included in previous sufficient conditions.

19.3.4 Smoothness of Constrained Optima

It's not much harder to show the solution to optimization problems with equality constraints are smooth.

Consider the problem with ℓ equality constraints and one parameter, a.

$$\begin{aligned} &\max_{\mathbf{x}} f(\mathbf{x}, a) \\ &\text{s.t } \mathbf{h}(\mathbf{x}, a) = \mathbf{0} \end{aligned} \tag{19.3.12}$$

where $f\colon U \to \mathbb{R}$ and $\mathbf{h}\colon V \times W \to \mathbb{R}^\ell$ with $U, V \subset \mathbb{R}^m$ and $W \subset \mathbb{R}$.

If the NDCQ holds, $\operatorname{rank} D_x\mathbf{h} = \ell$ at the optimal $\mathbf{x}^*(a)$.

The Lagrangian is

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, a) = f(\mathbf{x}, a) - \boldsymbol{\mu}\mathbf{h}(\mathbf{x}, a)$$

and any constrained optimum, $\mathbf{x}^*(a)$ satisfies the first order conditions

$$D_x\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, a) = \mathbf{0}, \quad D_\mu\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, a) = \mathbf{0}.$$

This is a system of $m + \ell$ equations (m from the first group, ℓ from the second). There are also $m + \ell$ endogenous variables $\mathbf{x}$ and $\boldsymbol{\mu}$, along with one exogenous variable a. Provided we know the problem has a solution (e.g., by Weierstrass's Theorem) the Implicit Function Theorem tells us the solution is $\mathcal{C}^1$ if the derivative of this system with respect to the endogenous variables is non-singular.

19.3.5 Implicit Functions and Constraint Qualification

We compute

$$D^2_{(\mu,x)}\mathcal{L} = \begin{pmatrix} 0_{\ell\times\ell} & -D\mathbf{h} \\ \\ -D\mathbf{h}^T & D^2_{\mathbf{x}}\mathcal{L} \end{pmatrix}. \tag{19.3.13}$$

It is the version of the bordered Hessian we previously encountered following Theorem 19.2.1. This matrix will be singular if NDCQ fails. That means that if we require $D^2_{(\mu,x)}\mathcal{L}$ be invertible, we are also requiring that the NDCQ is satisfied, as shown by the following theorem.

Theorem 19.3.2. *Suppose* $D^2_{(\mu,x)}\mathcal{L}(\mathbf{x}^*)$ *given by equation (19.3.13) is invertible. Then* $\operatorname{rank} D\mathbf{h}(\mathbf{x}^*) = \ell$.

Proof. Consider

$$\left[D^2_{(\mu,x)}\mathcal{L}(\mathbf{x}^*)\right]\begin{pmatrix}\mathbf{u}\\ \mathbf{v}\end{pmatrix} = \begin{pmatrix} -\big(D\mathbf{h}(\mathbf{x}^*)\big)\mathbf{v} \\ -\big(D\mathbf{h}(\mathbf{x}^*)\big)^T\mathbf{u} + \left[D^2_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*)\right]\mathbf{v}\end{pmatrix}$$

for any $\mathbf{u} \in \mathbb{R}^k$ and $\mathbf{v} \in \mathbb{R}^m$. If $D^2_{(\mu,x)}\mathcal{L}$ is invertible, $\ell + m \le \operatorname{rank} D\mathbf{h}(\mathbf{x}^*) + m \le \ell + m$, so we must have $\operatorname{rank} D\mathbf{h}(\mathbf{x}^*) = \ell$. ∎

19.3.6 Smooth Parameterization Theorem

The above reasoning leads to the following theorem.

Theorem 19.3.3. *Let* $\mathbf{x}^*(a)$ *be the solution to the parameterized constrained maximization problem in equation (19.3.12) and let* $\boldsymbol{\mu}^*(a)$ *be the corresponding vector of Lagrange multipliers. Fix* a_0*. If the bordered Hessian matrix* $D^2_{(\boldsymbol{\mu},\mathbf{x})}\mathcal{L}$ *is non-singular at* $\big(\mathbf{x}^*(a_0), \boldsymbol{\mu}^*(a_0), a_0\big)$*, then*

(a) *the NDCQ holds at* $\big(\mathbf{x}^*(a_0), \boldsymbol{\mu}^*(a_0), a_0\big)$*, and*

(b) $\mathbf{x}^*(a)$ *and* $\boldsymbol{\mu}^*(a)$ *are* $\mathcal{C}^1$ *at* $a = a_0$.

This theorem can also be generalized to a vector of parameters. Cases with inequality constraints are trickier. For the latter, it's necessary to focus on the binding constraints. Rather than considering a collection of variations on a theorem, we return to the issue of constraint qualification.

19.4 Constraint Qualification

The example that began in section 18.33 showed us that if constraint qualification fails, we may not be able to find the optimum.

In such cases we can try alternative methods of constraint qualification. There's even a theorem due to Fritz John where we can replace constraint qualification with an additional multiplier for the objective.

19.4.1 A Multiplier for the Objective

One alternative is to use an additional multiplier for the objective.

Theorem 19.4.1. *Let* f *and* h *be* $\mathcal{C}^1$ *functions on* $U \subset \mathbb{R}^2$. *Suppose that* $\mathbf{x}^*$ *solves*

$$\max_{\mathbf{x}} f(\mathbf{x})$$
$$s.t\ h(\mathbf{x}) = c$$

Form the Lagrangian

$$\mathcal{L}(\mathbf{x}, \mu_0, \mu_1) = \mu_0 f(\mathbf{x}) - \mu_1 \big(h(\mathbf{x}) - c\big),$$

by including a multiplier for the objective function f. *Then there exist multipliers* μ_0^* *and* μ_1^* *such that*

(a) μ_0^* *is either* 0 *or* 1,

(b) μ_0^* *and* μ_1^* *are not both zero,*

(c) $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ *satisfies the first order conditions*

$$\begin{aligned} \mathbf{0} &= D_{\mathbf{x}}\mathcal{L} = \mu_0 D_{\mathbf{x}} f - \mu_1 D_{\mathbf{x}} h \\ 0 &= \frac{\partial \mathcal{L}}{\partial \mu_1} = c - h(\mathbf{x}). \end{aligned}$$

Proof. If $D_{\mathbf{x}} h \neq \mathbf{0}$, constraint qualification (NDCQ) is satisfied, and there is a solution with $\mu_0^* = 1$ by Theorem 18.11.1.

If $D_{\mathbf{x}} h = \mathbf{0}$, then set $\mu_0^* = 0$. ∎

In this case, the extra multiplier μ_0 serves as a device to remind us to check the critical points of the constraints as well as the objective.

19.4.2 Example 18.33 Revisited I

Recall that the problem

$$\max_{(x,y)} x$$
$$\text{s.t. } x^3 + y^2 = 0.$$

showed us that without NDCQ constraint qualification, the standard first order conditions might not be necessary for an optimum.

Let's try Theorem 19.4.1 on it. Set up the modified Lagrangian

$$\mathcal{L} = \mu_0 x - \mu_1(x^3 + y^2).$$

The first order conditions based on the modified Lagrangian are

$$\begin{aligned} \mu_0 &= 3\mu_1 x^2 \\ 0 &= -2\mu_1 y \\ 0 &= x^3 + y^2. \end{aligned}$$

We saw in section 18.33 that when $\mu_0 = 1$, there is no solution. What if $\mu_0 = 0$?. Then the first of the first order condition becomes

$$0 = 3\mu_1 x^2,$$

so either $x = 0$ or $\mu_1 = 0$.

19.4.3 Example 18.33 Revisited II

If $x < 0$, $\mu_1 = 0$, which is forbidden by the condition that at least one μ_i must be non-zero.

Finally, if $x = 0$, $y = 0$, and any value of μ_1 will satisfy the first order equations. This yields the correct solution $(x, y) = (0, 0)$.

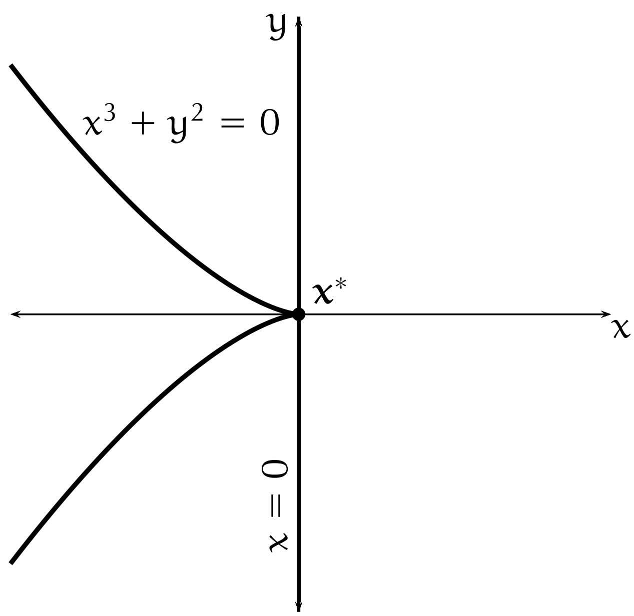

Figure 19.4.2: The solution is at the cusp, $\boldsymbol{x}^* = (0, 0)$. There $Df(\boldsymbol{x}^*) = (0, 0)$, forcing the multiplier $\mu_0 = 0$ because $Df = (1, 0) \neq (0, 0)$.

19.4.4 Fritz John Theorem

This version of Theorem 19.4.1, including a multiplier on the objective, allows for inequality constraints.[7]

Theorem 19.4.3. *Let* $U \subset \mathbb{R}^m$. *Suppose* $f\colon U \to \mathbb{R}$ *and* $\mathbf{g}\colon U \to \mathbb{R}^k$ *are* $\mathcal{C}^1$ *functions. Suppose further that* $\mathbf{x}^*$ *is a local max of* f *under the constraints*

$$\mathbf{g}(\mathbf{x}) \le \mathbf{b}.$$

Form the Lagrangian

$$\mathcal{L}(\mathbf{x}, \lambda_0, \ldots, \lambda_k) = \lambda_0 f(\mathbf{x}) - \sum_{j=1}^{k} \lambda_j \big(g_j(\mathbf{x}) - b_j\big)$$

with a multiplier λ_0 *for the objective function. Then there exist* $\lambda_0^*, \ldots, \lambda_k^*$ *such that:*

1. $D_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*, \lambda_0^*, \ldots, \lambda_k^*) = \mathbf{0}$.
2. $\lambda_1^*[g_1(\mathbf{x}^*) - b_1], \ldots, \lambda_k^*[g_k(\mathbf{x}^*) - b_k] = 0$.
3. $\lambda_1^* \ge 0, \ldots, \lambda_k^* \ge 0$.
4. $\mathbf{g}(\mathbf{x}^*) \le \mathbf{b}$.
5. $\lambda_0^* = 0$ *or* $\lambda_0^* = 1$, *and*
6. $(\lambda_0^*, \ldots, \lambda_k^*) \ne \mathbf{0}$.

[7] The work of Karush, Kuhn, Tucker, and John founded the subject of nonlinear programming, the study of nonlinear constrained optimization. Fritz John's original paper was Fritz John (1948), "Extremum problems with inequalities as subsidiary conditions" in K.O. Friedrichs et al. (eds.), Studies and Essays, Courant Anniversary Volume, Wiley/Interscience (Reprinted in: J. Moser (ed.): Fritz John Collected Papers vol. 2, Birkhäuser, 1985, pp. 543-560).

19.4.5 More Constraint Qualification: Maximia

There are other constraint qualification conditions that can be used. The following theorem collects some of the them without proof.

Theorem 19.4.4. *Let* $U \subset \mathbb{R}^m$ *and suppose* $f\colon U \to \mathbb{R}$ *and* $\mathbf{g}\colon U \to \mathbb{R}^k$ *are* $\mathcal{C}^1$ *functions. Suppose also that* $\mathbf{x}^*$ *is a local maximum of* f *under the constraints*

$$\mathbf{g}(\mathbf{x}) \leq \mathbf{b}.$$

Let $\hat{\mathbf{g}}(\mathbf{x}) \in \mathbb{R}^h$ *denote the vector of binding constraints at* $\mathbf{x}^*$*, and suppose* $\hat{\mathbf{g}}$ *obeys any of the following:*

(a) **NDCQ**: *The derivative* $D\hat{\mathbf{g}}(\mathbf{x}^*)$ *has rank* h.

(b) **Karush-Kuhn-Tucker CQ**: *For any* $\mathbf{v} \in \mathbb{R}^m$ *obeying* $D\hat{\mathbf{g}}(\mathbf{x}^*)\mathbf{v} \leq \mathbf{0}$ *there is an* $\varepsilon > 0$ *and a* $\mathcal{C}^1$ *curve* $\alpha\colon [0,\varepsilon) \to \mathbb{R}^m$ *such that:*
   (i) $\alpha(0) = \mathbf{x}^*$,
   (ii) $\alpha'(0) = \mathbf{v}$, *and*
   (iii) $g_i\big(\alpha(t)\big) \leq b_i$ *for* $i = 1, \dots, k$ *and all* $t \in [0, \varepsilon)$.

(c) **Slater CQ**: *There is a ball* $V \subset \mathbb{R}^m$ *containing* $\mathbf{x}^*$ *with the* g_i *convex on* V *and there exists a point* $\mathbf{z} \in V$ *with* $\mathbf{g}(\mathbf{z}) \ll \mathbf{b}$.

(d) **Concave CQ**: *The* g_i *are concave functions.*

(e) **Linear CQ**: *The* g_i *are linear functions.*

Then we can set $\lambda_0^* = 1$ *in Theorem 19.4.3.*

The first constraint qualification condition is our old friend, the NDCQ. The second was used by Kuhn and Tucker in their original work on optimization with inequality constraints. The Slater condition applies with convex constraints and requires that the constraint set have a non-empty interior. The concave CQ ensures the constraint set is convex. The last condition derives from linear programming models. It can be considered a special case of the Concave CQ.

19.4.6 More Constraint Qualification: Minima

Some of the constraint qualification conditions work a bit differently for minima. The way the constraints are written has been reversed in the following theorem. Further, the KKTCQ, Slater CQ, and Concave (now Convex) CQ are all modified.

Theorem 19.4.5. *Let* $U \subset \mathbb{R}^m$ *and suppose* $f\colon U \to \mathbb{R}$ *and* $\mathbf{g}\colon U \to \mathbb{R}^k$ *are* $\mathcal{C}^1$ *functions. Suppose also that* $\mathbf{x}^*$ *is a local minimum of* f *under the constraints*

$$\mathbf{g}(\mathbf{x}) \geq \mathbf{b}.$$

Let $\hat{\mathbf{g}}(\mathbf{x}) \in \mathbb{R}^h$ *denote the vector of binding constraints at* $\mathbf{x}^*$*, and suppose* $\hat{\mathbf{g}}$ *obeys any of the following:*

(a) **NDCQ**: *The derivative* $D\hat{\mathbf{g}}(\mathbf{x}^*)$ *has rank* h.

(b) **Karush-Kuhn-Tucker CQ**: *For any* $\mathbf{v} \in \mathbb{R}^m$ *obeying* $D\hat{\mathbf{g}}(\mathbf{x}^*)\mathbf{v} \geq \mathbf{0}$ *there is an* $\varepsilon > 0$ *and a* $\mathcal{C}^1$ *curve* $\alpha\colon [0,\varepsilon) \to \mathbb{R}^m$ *such that:*
   - (i) $\alpha(0) = \mathbf{x}^*$,
   - (ii) $\alpha'(0) = \mathbf{v}$, *and*
   - (iii) $g_i\big(\alpha(t)\big) \geq b_i$ *for* $i = 1, \ldots, k$ *and all* $t \in [0,\varepsilon)$.

(c) **Slater CQ**: *There is a ball* $V \subset \mathbb{R}^m$ *containing* $\mathbf{x}^*$ *with each* g_i *concave on* V *and there exists a point* $\mathbf{z} \in V$ *with* $\mathbf{g}(\mathbf{z}) \gg \mathbf{b}$.

(d) **Convex CQ**: *The* g_i *are convex functions.*

(e) **Linear CQ**: *The* g_i *are linear functions.*

Then we can set $\lambda_0^* = 1$ *in Theorem 19.4.3.*

December 6, 2022