Chapter 5

Optimization

5.1 Optimization

Virtually all of microeconomics is predicated upon the idea that the (economic) behavior of people can be explained as an attempt on their part to maximize (or minimize) something subject to some constraints. This view does not necessarily ascribe much intelligence to the people whose behavior is to be described — the behavior of a rock falling off a cliff can be described, after all, as an attempt on the part of the rock to minimize its potential energy. It simply means that we microeconomic theorists always think we can cook up models based on optimizing behavior which will capture the important aspects of actual behavior.

5.1.1 Unconstrained Optimization

The simplest optimization problem involves no constraints. Here $f : \mathbb{R}^n \mapsto \mathbb{R}$ is the *objective function* and the objective, accordingly, is to find that x^* which makes $f(x^*)$ as large (or as small) as possible:

$$x^* \equiv \begin{cases} \arg\ \max_x f(x) & \text{if} \quad f(x^* + v) - f(x^*) \leq 0,\ \forall v \\ \arg\ \min_x f(x) & \text{if} \quad f(x^* + v) - f(x^*) \geq 0,\ \forall v \end{cases}$$

Here the terms *arg max* and *arg min* refer to the "arguments" which solve the respective optimization problem.

Now if f is differentiable and $f_x(x^*)$ is its gradient evaluated at the point x^* then

$$f(x^* + v) - f(x^*) \approx f_x(x^*)v$$

Thus $f_x(x^*) = 0$ is a necessary condition either for a maximum or a minimum — if it were not zero then f would increase for some choice of $\hat{v}$ (not a maximum) and fall for $-\hat{v}$ (not a minimum either).

The requirement that the gradient "vanish", $f_x(x^*) = 0$, entails solving the system of n equations:

$$\begin{aligned} \frac{\partial f}{\partial x_1}(x_1^*, \ldots, x_n^*) &= 0 \\ \frac{\partial f}{\partial x_2}(x_1^*, \ldots, x_n^*) &= 0 \\ &\vdots \\ \frac{\partial f}{\partial x_n}(x_1^*, \ldots, x_n^*) &= 0 \end{aligned}$$

for the n unknown components of x^*. Since this requirement involves the first differential and is a necessary condition it is called the *first order necessary condition*.

Another necessary condition involves the second differential. If f is twice differentiable with second derivative $f_{xx}(x)$

$$\begin{aligned} f(x^* + v) - f(x^*) &\approx f_x(x^*)v + \frac{1}{2}v^T f_{xx}(x^*)v \\ &\approx \frac{1}{2}v^T f_{xx}(x^*)v \text{ when } f_x(x^*) = 0 \end{aligned}$$

Moreover, since

$$f(x^* + v) - f(x^*) \quad \begin{cases} \leq 0 & \text{if } f_{xx}(x^*) \text{ is negative semidefinite} \\ \geq 0 & \text{if } f_{xx}(x^*) \text{ is positive semidefinite} \end{cases}$$

we have the *second order necessary condition*:

Theorem 25. For x^* to impart a maximum (minimum) to $f(x)$ it must be the case that $f_{xx}(x^*)$ is negative (positive) semidefinite.

It is sufficient that $f_{xx}(x^*)$ be negative definite for a maximum or positive definite for a minimum, but *sufficient conditions* aren't very important in Microeconomics. The reason is that Economic models typically involve the optimization hypothesis that the person whose behavior is being modeled behaves as if he/she were trying to maximize (or minimize) an objective function subject to constraints. Call this optimizing model M. A necessary condition, N, which must be true in order for M to be true provides a *testable hypothesis* — see Figure 5.1 — if evidence suggests that N is false then M must be false as well. Sufficient conditions, on the other hand, do not provide testable hypotheses — S, for example, is sufficient for M in Figure 5.1 but evidence that S is false does not imply that M is false.

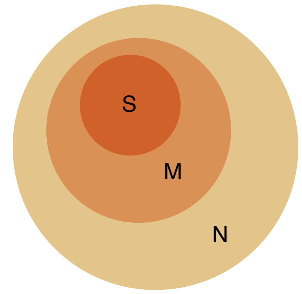

Figure 5.1: Necessary and Sufficient Conditions for M

□ *Problem* 5.1. In *Mathematica* the command `Maximize[f,{x,y,...}]` maximizes f with respect to x, y, ... Describe the output from

```
f = -(x + 2) (x^2 + 1) x (x - 1) (x - 2);
Plot[f,{x,-2.5,2.5}]
Maximize[f,x]//N
```

5.1.2 Constrained Optimization

The analogous constrained optimization problem is

$$x^* = \arg\max_x f(x)$$
$$\text{s.t.} \quad g^i(x) \geq 0, \quad i = 1, \ldots, m$$

The requirements that x must satisfy $g^i(x) \geq 0$, $i = 1, \ldots, m$, are called *constraints* and the problem is to find that x^* which satisfies the constraints and which imparts the largest value to the objective function, i.e., x^* solves this problem if and only if

- $g(x^*) \geq 0$
- for any $\hat{x}$, $f(\hat{x}) > f(x^*)$ implies $g(\hat{x}) \ngeq 0$.

There are exactly two possibilities for the solution to this problem, x^*,

- There is no $\hat{x}$ for which $f(\hat{x}) > f(x^*)$. Here the constraints are "not binding" — can be erased without affecting the solution to the problem. Here the unconstrained conditions hold.
- There is a $\hat{x}$ for which $f(\hat{x}) > f(x^*)$ but, of course, $g(\hat{x}) \ngeq 0$. Here the constraints are binding — prevent one from doing as well as would be possible without them — and the unconstrained conditions do not necessarily hold.

Focusing upon the second possibility for the moment, notice that if the objective function, f, and the *binding* constraints, $\hat{g}$, are differentiable at x^* then

$$f(x^* + v) - f(x^*) \approx f_x(x^*)v$$

and

$$\begin{aligned}\hat{g}(x^* + v) - \hat{g}(x^*) &= \hat{g}(x^* + v)\\ &\approx \hat{g}_x(x^*)v\end{aligned}$$

where $\hat{g}_x(x^*)$ denotes the matrix whose rows are the gradients of those constraints which are *binding* at x^*. For a maximum it is intuitive that the inequalities:

$$\hat{g}_x(x^*)v \geq 0, \quad (-f_x)v < 0 \tag{5.1}$$

cannot have any solutions since:

1. For small v, the movement from x^* to $x^* + v$ would satisfy the constraints. This follows from the fact that the components of $\hat{g}(x^*)$ are all equal to zero, all other constraints are strictly positive and the (small) movement to $x^* + v$ would thus leave all constraints non-negative.
2. For small v, the movement from x^* to $x^* + v$ would impart a larger value to the objective function.

This intuition is basically correct save for pathological cases — see Arrow, Hurwitz and Uzawa for a discussion of *constraint qualifications* which are sufficient to eliminate such problems. [The original Kuhn-Tucker constraint qualification was that the gradients of the binding constraints must be linearly independent.]

Recall now that Farkas' Lemma (Theorem 6 on page 31) says that if A is an m by n matrix and $b \neq 0$ is a 1 by n row vector then exactly one of the following is true:

1. $yA = b, \ y > 0$ has a solution $y \in \mathbb{R}^m$
2. $Az \geq 0, \ b \cdot z < 0$ has a solution $z \in \mathbb{R}^n$

Now make the association that $A = \hat{g}_x(x^*)$, $b = -f_x(x^*)$, $z = v$, $y = \lambda$ and note that $b = -f_x(x^*) \neq 0$ by virtue of the fact that we are focusing upon the case in which the unconstrained, vanishing gradient, condition does not hold. Then either

$$\lambda \hat{g}_x(x^*) = -f_x(x^*), \quad \lambda > 0 \tag{5.2}$$

has a solution or Equation 5.1 has a solution. But since there can be no solutions to Equation 5.1 if x^* is to solve the maximization problem, it follows from Farkas' Lemma that Equation Equation 5.2 must have a solution.

Equation 5.2 is illustrated for a simple maximization problem in Figure 5.2 on the next page. Here there are three constraints and the problem is to

$$\begin{aligned} &\max_x f(x) = (1,1)\cdot(x_1,x_2)\\ \text{s.t.}\quad & g^1(x) = (1,0)\cdot(x_1,x_2) \geq 0\\ & g^2(x) = (0,1)\cdot(x_1,x_2) \geq 0\\ & g^3(x) = 1 + (-1/3,-1)\cdot(x_1,x_2) \geq 0 \end{aligned} \tag{5.3}$$

This is an example of a *linear programming problem* — a problem in which both the objective function and the constraints are linear.

Note first that the shaded area corresponds to $g(x) \geq 0$ — the set of x's which satisfy all three constraints. By inspection it can be seen that $x^* = (3,0)$ solves the problem since:

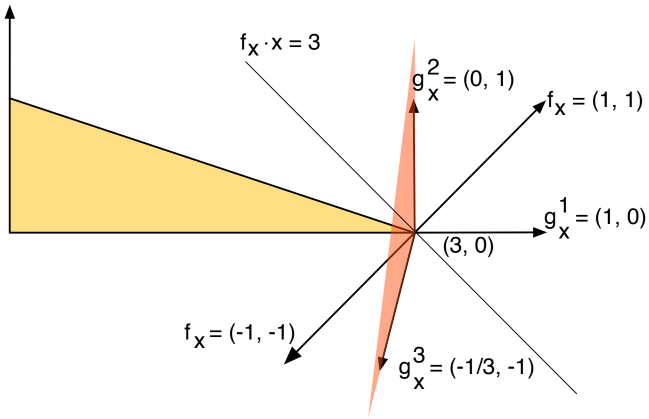

Figure 5.2: Two Binding Constraints

- $(3, 0)$ belongs to the shaded area
- $(3, 0)$ imparts a value of 3 to the objective function
- x's which make the objective function larger than 3 do not belong to the shaded area.

That this point satisfies Equation 5.2 on the previous page can be seen by plotting, with their tails at $(3, 0)$, the gradients of the objective function and each of the three constraints. Note that $-f_x$ belongs to the cone generated by g^2 and g^3, the gradients of the constraints which are actually binding at $(3, 0)$, since it is possible to express $-f_x$ as a positive linear combination of g^2 and g^3. These positive weights provide the values of the components of λ in the solution to $\lambda \hat{g}_x = -f_x$ with $\lambda > 0$:

$$\begin{bmatrix} \lambda_2 & \lambda_3 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1/3 & -1 \end{bmatrix} = (-1, -1)$$

or

$$\begin{bmatrix} \lambda_2 & \lambda_3 \end{bmatrix} = \begin{bmatrix} 2 & 3 \end{bmatrix}$$

□ *Problem* 5.2. For the problem of Figure 5.2 show geometrically why the point $(0, 1)$ does not satisfy Equation 5.2 on the previous page. [Recall that $\hat{g}_x(0, 1)$ contains only the gradients of the constraints which are binding at $(0, 1)$.]

If we write the *Lagrangian Function*

$$L(x, \lambda) \equiv f(x) + \sum_{i=1}^{m} \lambda_i g^i(x)$$

then, with slight modification, the "Farkas' Lemma" requirement for a maximum, Equation 5.2 on the previous page, becomes the *Kuhn-Tucker Conditions* which, given the constraint qualification, are necessary for a solution to the maximization problem:

$$L_x(x^*, \lambda^*) = 0 \tag{5.4}$$
$$\lambda L_\lambda(x^*, \lambda^*) = 0 \tag{5.5}$$
$$L_\lambda(x^*, \lambda^*) \geq 0 \tag{5.6}$$
$$\lambda^* \geq 0 \tag{5.7}$$

Equation 5.4 on preceding page, Equation 5.5 on preceding page and Equation 5.7 on preceding page are the solution to Equation 5.2 on page 60 save for the fact that λ is now allowed to be equal to zero — this to incorporate the case in which none of the constraints is binding. Equation 5.6 on preceding page is merely a restatement of the constraints. Equation 5.5 on preceding page and Equation 5.7 on preceding page form the so-called *complementary slackness condition* which prevents non-binding constraints from playing a role in Equation 5.4 on preceding page. Equation 5.5 on preceding page says that

$$\sum_{i=1}^{m} \lambda_i^* g^i(x^*) = 0$$

Since each term in this summation is the product of two non-negative numbers (Equation 5.6 on preceding page and Equation 5.7 on preceding page), it follows that $g^i(x^*) > 0$ implies $\lambda_i^* = 0$ and, conversely, $\lambda_j^* > 0$ implies $g^j(x^*) = 0$. If the ith constraint is not binding, $g^i(x^*) > 0$ then the associated *Lagrangian multiplier*, λ_i^*, must be equal to zero. Conversely, if $\lambda_i^* > 0$ then the corresponding constraint must be binding, $g^i(x^*) = 0$. Thus Equation 5.4, Equation 5.5 and Equation 5.7 on preceding page are equivalent to Equation 5.2 on page 60.

□ *Problem* 5.3. Is it possible for both $\lambda_j^* = 0$ and $g^j(x^*) = 0$ to hold in an optimal solution? If so, provide an example of an optimization problem in which this would be true. Otherwise show why this is not possible.

The complementary slackness condition is consistent with the *shadow price* interpretation of the Lagrangian multipliers in which λ_i is the value of relaxing the ith constraint. This can expressed more precisely as follows. Introduce a vector of *parameters*, $a = (a_1, \ldots, a_m)$ into the generic optimization problem as follows

$$\begin{aligned} x^*(a) \equiv \arg\max_x f(x) & \\ \text{s.t. } & g^1(x) + a_1 \geq 0 \\ & \vdots \\ & g^m(x) + a_m \geq 0 \end{aligned}$$

The presence of the a's in the constraints means that the optimal solution(s) will depend upon which values are selected for these a's. The notation $x^*(a)$ reflects this functional dependence — it is simply the set of x's which solve the problem for the given a. Now the value of the objection function also depends upon a. This can be expressed by

$$\begin{aligned} F(a) \equiv \max_x f(x) & \\ \text{s.t. } & g^1(x) + a_1 \geq 0 \\ & \vdots \\ & g^m(x) + a_m \geq 0 \end{aligned}$$

Notice that $F(a)$, unlike $x^*(a)$, is necessarily a single valued function — there can be but one value for the maximized objective function. [In the previous example $F(a) = a$.] Provided that it is differentiable then

$$\lambda_i = \frac{\partial F(0)}{\partial a_i} \tag{5.8}$$

represents the rate at which the (optimized) value of the objective function increases as the ith constraint is relaxed or as a_i is increased from zero.

□ *Problem* 5.4. In *Mathematica* the command `Maximize[{f,cons},{x,y,...}]` maximizes `f` with respect to `x,y,...` subject to the constraints, `cons`. How well does *Mathematica* do with the illustrative problem in Equation 5.3 on page 60?

```
Maximize[{x1+x2, x1>=0, x2>=0, 1-x1/3-x2>=0},{x1,x2}]
```

5.1.3 Roots of Symmetric Matrices

The Kuhn-Tucker first-order conditions can themselves be used to derive some results related to the second-order conditions.

Theorem 26. If A is symmetric and if either

$$\begin{aligned} x' &= \arg\max_x \; x^T A x \\ &\text{s.t. } x^T x \le 1 \\ \lambda' &= \max_x \; x^T A x \\ &\text{s.t. } x^T x \le 1 \end{aligned}$$

or

$$\begin{aligned} x' &= \arg\min_x \; x^T A x \\ &\text{s.t. } \; x^T x \ge 1 \\ \lambda' &= \min_x \; x^T A x \\ &\text{s.t. } \; x^T x \ge 1 \end{aligned}$$

then x' is a characteristic vector of A and λ' is the associated characteristic root.

□ *Problem* 5.5. Show, for any symmetric n by n matrix A, that the gradient of the associated quadratic form is given by

$$\frac{\partial x^T A x}{\partial x} = 2Ax$$

For the maximization problem we have the Lagrangian function

$$L_{\max} = x^T A x + \lambda[1 - x^T x]$$

while for the minimization problem we maximize $-x^T A x$ and have

$$L_{\min} = -x^T A x + \lambda[x^T x - 1]$$

Thus

$$-L_{\min} = L_{max}$$

and in either case a necessary condition is that

$$L_x = 2Ax - 2\lambda x = 0$$

or

$$Ax = \lambda x$$

Since an optimal x must satisfy the characteristic equation it follows that an optimal x must be a characteristic vector of A and the Lagrangian multiplier λ must be the associated characteristic root. Pre-multiply both sides of the last expression by x^T to obtain

$$\begin{aligned} x^T A x &= \lambda x^T x \\ &= \lambda 1 \\ &= \lambda \end{aligned}$$

Thus the Lagrangian multiplier/characteristic root is, in fact, equal to the optimized value of the objective function.

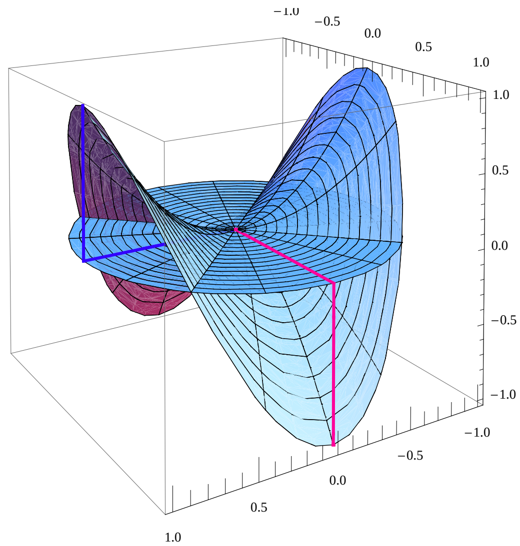

Figure 5.3: Characteristic Roots and Vectors

Theorem 26 on preceding page is illustrated in Figure 5.3 for the case in which

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

The characteristic roots and associated characteristic vectors for this indefinite quadratic form are

$$\begin{aligned} \lambda_1 &= 1; \quad x^1 = (1, 0) \text{ or } (-1, 0) \\ \lambda_2 &= -1; \quad x^2 = (0, 1) \text{ or } (0, -1) \end{aligned}$$

Moving around the unit circle, $x^T x = 1$, the point x^1 (x^2) corresponds to the point at which $x^T A x$ is furthest above (below) the x-plane and λ_1 (λ_2) is the actual distance from the x-plane to the graph $x^T A x$ at this point.

Theorem 27. It is necessary and sufficient for the symmetric matrix A to be

1. a positive definite quadratic form that all the characteristic roots of A be positive.
2. a positive semi-definite quadratic form that all the characteristic roots of A be non-negative.
3. an indefinite quadratic form that at least one of the characteristic roots of A be positive and at least one be negative.

Note that it is sufficient to explore the "unit circle" $\{x \in \mathbb{R}^n \mid x^T x = 1\}$ since the sign of the quadratic form at an arbitrary point $x \neq 0$ must be the same as the sign at the point $x/|x|$

$$\frac{x^T}{|x|} A \frac{x}{|x|} = \frac{1}{|x|^2} x^T A x, \quad \forall x \neq 0$$

and the latter point is on the unit circle. Suppose, for example, that A is positive definite. Then since $x^T A x > 0$ for all $x \neq 0$ it follows that

$$\begin{aligned} \underline{\lambda} &\equiv \min_x x^T A x \\ &\quad \text{s.t. } x^T x \geq 1 \\ &> 0 \end{aligned}$$

In view of Theorem 26 on page 63, if $\underline{\lambda}$ must be the smallest of all of the characteristic roots it follows that all of the roots must be positive. Conversely, if all of the roots are positive then the minimization problem has a strictly positive solution and the quadratic form is positive definite. A similar argument holds if the quadratic form is positive semi-definite, save for the fact that the smallest root may actually be equal to zero. Finally, if the quadratic form is indefinite then its minimum on the unit circle will clearly be negative and its maximum will be positive and there will, accordingly, be characteristic roots of both signs. Conversely, if there are both positive and negative characteristic roots then the quadratic form takes on both positive and negative values and is therefore indefinite.

Finally, in view of Theorem Theorem 25 on page 58:

Theorem 28. Suppose $f : \mathbb{R}^n \mapsto \mathbb{R}$ is continuously twice differentiable. For x^* to impart a maximum (minimum) to $f(x)$ it must be the case that the characteristic roots of $f_{xx}(x^*)$ are non-positive (non-negative).

□ *Problem* 5.6. Consider the (unconstrained) problem of minimizing with respect to `x` and `y` the function

```
f = x^4 + 3 x^2 y + 5 y^2 + x + y;
```

Using *Mathematica*, first plot this function, then solve the minimization problem and finally check the solution by examining the gradient and the characteristic roots (eigenvalues) of the hessian.

```
Plot3D[f, {x, -3, 3}, {y, -6, 6}]
min = Minimize[f, {x, y}] // N
sol = Last[min]
gradient[f_, vars_List] := Map[D[f, #] &, vars]
hessian[f_, vars_List] := Outer[D, gradient[f, vars], vars]
gradient[f, {x, y}] /. sol
hes = hessian[f, {x, y}] /. sol
Eigenvalues[hes]
```

Comment on the output and indicate whether or not *Mathematica*'s solution satisfies the first and second order conditions for a minimum.

□ *Problem* 5.7. Repeat Problem 5.6 this time for the function

```
f = x^3 + y^3 + 2 x^2 + 4 y^2 + 6
```

and using `FindMaximum[f, {{x, 0}, {y, -8/3}}]` to find a *local maximum* of `f` starting at `x = 0` and `y = -8/3`, this time restricting the plot range to `{x,-2.5,1}, {y,-3.5,1}`. Are any warnings issued by *Mathematica* regarding the solution? Is the solution actually a local maximum, a local minimum or a local saddle point?

□ *Problem* 5.8. Consider a "knapsack problem" (integer programming problem) in which five items are available for packing in the knapsack. Let the integer $x_i = \{0, 1\}$ denote whether (1) or not (0) the ith item is included. The benefit to including the items are, respectively, 14, 10, 15, 8 and 9 and the corresponding weights are 6, 8, 5, 6 and 4. Which items should be included if the goal is to maximize the total benefit subject to the constraint that the total weight not exceed 18? Hint: Use *Mathematica*:

```
Maximize[{14x1+10x2+15x3+8x4+9x5,
6x1+8x2+5x3+6x4+4x5 <= 18,
0<=x1<=1, 0<=x2<=1, 0<=x3<=1, 0<=x4<=1, 0<=x5<=1,
Element[{x1,x2,x3,x4,x5}, Integers]},
{x1,x2,x3,x4,x5}]
```

□ *Problem* 5.9. *Mathematica* uses the optimized command `LinearProgramming[c,m,b]` to solve the *linear programming* problem of minimizing $c \cdot x$ subject to the constraints that $mx \geq b$ and $x \geq 0$ where c and x are n-tuples, m is an $m \times n$ matrix and b is an m-tuple. Use it to solve the "transportation problem" in which there are three sources of supply with quantities 47, 36 and 52, respectively, and four destinations with demands 38, 34, 29 and 34, respectively. In this problem x will be the 12-tuple `x11,...,x34` where `xIJ` denotes the quantity shipped from source `I` to destination `J`. The 12-tuple of the corresponding costs, `cIJ`, of shipping one unit from source `I` to destination `J` is

```
c = {5, 7, 6, 10, 9, 4, 6, 7, 5, 8, 6, 6}
```

The problem is to minimize the total shipping cost of meeting the specified demands from the available sources. There are seven constraints, three specifying that total shipments from each of the three sources cannot exceed the available amounts and another four specifying that the total shipments to each of the four destinations cannot be less than the demands. To express the constraints in the required $\geq$ format we then need, for example, the first row of `m` to be

```
{-1, -1, -1, -1, 0, 0, 0, 0, 0, 0, 0, 0}
```

and the first component of `b` to be -47. Complete the specification of `m` and `b` and give the solution from `LinearProgramming[c,m,b]`. What is the total shipping cost in the optimal solution? How many of the constraints are binding in the optimal solution?

5.1.4 The Parametric Problem

The typical "optimizing model of behavior" in microeconomics adds one final ingredient — parameters — to the constrained optimization problem:

$$
\begin{aligned}
x(a) &\equiv \arg\max_x f(x,a) \\
&\text{s.t. } g^i(x,a) \geq 0, \quad i = 1, \ldots, mF(a) \equiv \max_x \qquad f(x,a) \\
&\text{s.t. } g^i(x,a) \geq 0, \quad i = 1, \ldots, m
\end{aligned}
\tag{5.9}
$$

where

$x \in \mathbb{R}^n$ the explained or endogenous variables
$a \in \mathbb{R}^p$ the parameters or exogenous variables
$x(a)$ the reduced form, a point-to-set mapping

Note that this is a *generic* optimization problem in the sense that it incorporates:

1. Minimization: replace min $h(x, a)$ with max $f(x, a)$ where $f(x, a) \equiv -h(x, a)$
2. Arbitrary inequality constraints: replace $w(x, a) \leq z(x, a)$ with $g(x, a) \equiv z(x, a) - w(x, a) \geq 0$
3. Equality constraints: replace $h(x, a) = 0$ with the two constraints $g^1(x, a) \equiv h(x, a) \geq 0$ and $g^2(x, a) \equiv -h(x, a) \geq 0$

The vector x represents the *endogenous variables* whose values will be chosen, according to the model, by some person as if to maximize the objective function subject to the constraints. The business of the model is to predict the values of these variables — hence the term "endogenous". The vector a represents the parameters or *exogenous variables* whose values are not explained by the model but rather are exogenously specified — the "dial settings" for the experiment. In consumption theory, for example, a would correspond to the consumer's income and the market prices of consumption goods and x to the quantities of these consumption goods demanded. The mapping, $x(a)$, would in this case be called the consumer's demand function.

The purpose of an optimizing model is to explain the choice of x in experiments which are characterized by the given values of a or, in short, to ascertain the properties of the *point-to-set mapping* $x(a)$ — the term "point-to-set" reflects the fact that the image of a may be a set rather than a single point. This mapping might be called the *reduced form* of the model. It is a sort of "bottom line" for the model since it expresses the endogenous or explained variables directly as functions of the exogenous or explanatory variables. To predict the result of an experiment simply stick the description of the experiment, a, into this function and out pops the set of possible results, $x(a)$.

Models can be more or less specific. In the typical example problem or exercise both the objective function and constraints are explicitly stated and it is possible to determine the exact properties of $x(a)$ since it is generally possible to solve for the reduced form explicitly. In most "real" analysis, on the other hand, neither the objective function nor the constraints are explicitly stated. Instead *qualitative properties* of these functions are given and the analytical challenge becomes to trace these qualitative properties through the optimization problem to discover what qualitative properties they entail for the reduced form.

A "conservation of information" law holds for this deductive process — the more (or less) specific the information about the objective function and constraints the more (or less) specific will be the derived information about the reduced form. In a model in which all functions are explicitly stated, for example, it might be possible to deduce that an increase in a given sales tax by five cents would reduce the given consumer's consumption of the taxed commodity by ten pints per week. In a less specific model, on the other hand, it might only be possible to state that an increase in the sales tax would entail a reduction (by some indeterminate amount) in the quantity consumed. Such results are called *comparative static effects* and are typically expressed as propositions concerning the *signs* of partial derivatives of the reduced form.

5.2 The Well Posed Problem

Certain potential problems of optimization models are best recognized and eliminated as early in the process as possible:

1. Is it possible for $x(a)$ to be empty, i.e, to contain no solutions whatever? Since this amounts to the prediction that nothing is possible, this would be a rather serious, indeed fatal, flaw in a model.
2. Is it possible for $x(a)$ to contain more than one solution? The limiting case here would be that anything is possible or that the model has no predictive value — it doesn't rule out any possibilities whatever.

3. Is it possible for a single valued $x(a)$ not to be smooth, i.e. discontinuous or non-differentiable? The "pre Chaos Theory" view of the world of actual experiments was that small changes in exogenous parameters produce correspondingly small changes in endogenous variables — that $X(A)$ is smooth. The "post Chaos Theory" view allows for "butterfly effects" — e.g. the possibility that the small change in air currents caused by a butterfly's wings could cause a large change in weather patterns. Whatever one's view of such matters, computing comparative static effects is certainly much easier when they can be obtained as (partial) derivatives.

How can such problems be fixed?

5.2.1 The Existence Problem

The optimization problem will have no solutions only for the most poorly posed problems. An suggestive example is the problem of finding the largest real number which is strictly less than one. There is no solution for this constrained maximization problem nor is there one for the unconstrained problem of simply choosing the largest real number. The failures in both problems stems from the same problem — the *feasible set*, the set of points which satisfy the constraint(s), is not compact.

Recall from Theorem 21 on page 45 that a continuous function must attain a maximum and a minimum on a compact set and from Theorem 19 on page 45 that a subset of $\mathbb{R}^n$ will be compact if and only if it is closed and bounded.

Note the failure of the feasible sets in the two examples. The set of real numbers strictly less than one is not closed and the set of all real numbers is not bounded. Compactness turns out to be just the right requirement for feasible sets in view of the following two related facts:

Stock assumptions which assure that a problem has at least one solution are thus (i) a continuous objective function and (ii) constraints which yield a compact feasible set.

5.2.2 The Uniqueness Problem

Now it would also be nice if $x(a)$ were a function and not a point-to-set mapping since its nicer to have a model which makes a unique prediction regarding behavior than one which merely enumerates a collection of possibilities. The predictive power of the model is greater and comparative statics results, which require functions, become possible. Unfortunately the "arguments" which solve maximization problems need not generally be unique. There are, however, special classes of problems for which the solutions are necessarily unique and, thus, for which $x(a)$ is a function. One of the most important of these involves problems whose objective functions and constraints are quasi-concave.

The connection of quasi-concavity to the uniqueness of solutions to Equation 5.9 on page 66 is immediate. Suppose, for example, that the constraints, $g^i(x,a)$, are (strictly) quasi-concave functions in x. Then the set of x's which satisfy $g^i(x,a) \geq 0$ is simply a level set for the (strictly) quasi-concave function and thus a (strictly) convex set. The feasible set, the set of x's which satisfy all of the constraints, as an intersection of (strictly) convex sets must itself be (strictly) convex.

Now add the fact that the objective function, $f(x,a)$, is quasi-concave in x and $x(a)$ must then be a convex set. This follows from two observations. If any two distinct points belong to $x(a)$ then (i) both points must belong to the feasible set and (ii) both points must impart the same value to the objective function. It follows from the first and the convexity of the feasible set that any point along the line segment connecting the two points must belong to the feasible set. It follows from the second and the quasi-concavity of the objective function that any point along this line segment must impart a value to the objective function which is at least as great as the value at the endpoints. But the value of the objective function at any point along the line segment also cannot exceed the value at the end points else the end points could not have belonged to $x(a)$. Hence the value of the objective function must be constant along the line segment and the entire line segment must then belong to $x(a)$.

If the constraints are quasi-concave and the objective function is *strictly* quasi-concave then $x(a)$ can contain at most a single point. To see this suppose to the contrary that $x, x' \in x(a)$ with $x \neq x'$ and pursue the argument of the previous paragraph. Any point along the line segment connecting x and x' must be feasible and must now impart a value to the objective function which is, because of strict quasi-concavity, *strictly* greater than either of the endpoints. Since this contradicts the assumption that $x \neq x' \in x(a)$ it cannot be the case that $x(a)$ contains two or more points — either $x(a)$ is empty or it contains a unique solution and can be regarded as a function.

5.2.3 The Differentiability Problem

Optimization models in which the mapping $x(a)$ is a not only a function but differentiable as well are particularly important in Economics. In such models comparative static effects are easily expressed as the partial derivatives of $x(a)$.

Suppose that in Equation 5.9 on page 66 the constraints are quasi-concave and the objective function is strictly quasi-concave in x and differentiable in both x and a. Then the problem has a unique solution (from strict quasi-concavity) which can be characterized by the Kuhn-Tucker Conditions (from differentiability in x). The Lagrangian Function is

$$L(x, \lambda, a) \equiv f(x, a) + \sum_{j=1}^{m} \lambda_j g^j(x, a)$$

and the Kuhn–Tucker necessary conditions are

$$\begin{aligned} L_x\left(x(a), \lambda(a), a\right) &= 0 \\ \lambda(a) \cdot L_\lambda\left(x(a), \lambda(a), a\right) &= 0 \\ L_\lambda\left(x(a), \lambda(a), a\right) &\geq 0 \\ \lambda(a) &\geq 0 \end{aligned}$$

Now choose values for the exogenous variables, $a = \hat{a}$, suppose the non-binding constraints have been eliminated and write the Kuhn-Tucker conditions as

$$\begin{aligned} L_x(\hat{x}, \hat{\lambda}, \hat{a}) &= 0 \\ g(\hat{x}, \hat{a}) &= 0 \end{aligned} \tag{5.10}$$

where

$$\begin{aligned} L(x, \lambda, \hat{a}) &\equiv f(x, \hat{a}) + \lambda g(x, \hat{a}) \\ g(x, \hat{a}) &\equiv [g^1(x, \hat{a}), \ldots, g^m(x, \hat{a})] \end{aligned}$$

When Equation 5.10 is solved for $x(\hat{a})$ will the result be differentiable? The Implicit Function Theorem provides the answer to this question.

Theorem 29 (Implicit Function Theorem). Suppose $h^i(y, a), i = 1, , \ldots, l$ are real-valued and continuously differentiable functions on $\mathbb{R}^l \times \mathbb{R}^p$ with $y \in Y$, an open subset of $\mathbb{R}^l$, and $a \in A$, an open subset of $\mathbb{R}^p$. Let $h^i(\hat{y}, \hat{a}) = 0,\ i = 1, \ldots, l$. Provided that the determinant of the Jacobian does not vanish, $|h_y(\hat{y}, \hat{a})| \neq 0$, there exists a *continuously differentiable* function $y(a)$ such that $\hat{y} = y(\hat{a})$ and $h^i(y(a), a) = 0,\ i = 1, \ldots, l$ for all a in some neighborhood of $\hat{a}$.

Intuitively, this theorem says that if you fix $a = \hat{a}$ you can regard h as a mapping which takes a point y in $\mathbb{R}^l$ and produces another point, $h(y, \hat{a})$, in $\mathbb{R}^l$. If this mapping is differentiable it has a good linear approximation given by the Jacobian h_y. If the determinant of this Jacobian is non-singular then the linear approximation is invertible. This means that it's possible to find out what y mapped into zero when $a = \hat{a}$, namely, $\hat{y} = y(\hat{a})$. The theorem states that it is not only possible to find $\hat{y}$, but, locally at least, that it is also possible to find the *function* $y(a)$ and that this function will be differentiable.

To use the Implicit Function Theorem, let the first order conditions of Equation 5.10 on preceding page correspond to the h^i, let l correspond to $n + m$ and let y correspond to (x, λ). Then under the conditions of the implicit function theorem there is a continuously differentiable function $y(a) = [x(a), \lambda(a)]$ such that

$$\begin{aligned} L_x(x(a), \lambda(a), a) &= 0 \\ g(x(a), a) &= 0 \end{aligned} \tag{5.11}$$

for all a in some neighborhood of $\hat{a}$.

5.3 Comparative Statics

5.3.1 The Classic Approach

Since the Implicit Function Theorem assures that $x(a)$ is differentiable, we can differentiate Equation 5.11 with respect to a_k to obtain

$$\begin{bmatrix} L_{xx} & g_x^T \\ g_x & 0 \end{bmatrix} \begin{bmatrix} \partial x / \partial a_k \\ \partial \lambda / \partial a_k \end{bmatrix} + \begin{bmatrix} L_{x a_k} \\ L_{\lambda a_k} \end{bmatrix} = 0$$

Since the matrix on the left side of this expression (the Jacobian of the h^i's) is assumed to be non-singular, the inverse exists and we have

$$\begin{bmatrix} \partial x / \partial a_k \\ \partial \lambda / \partial a_k \end{bmatrix} = - \begin{bmatrix} L_{xx} & g_x^T \\ g_x & 0 \end{bmatrix}^{-1} \begin{bmatrix} L_{x a_k} \\ L_{\lambda a_k} \end{bmatrix} \tag{5.12}$$

Equation 5.12 is called *fundamental equation of comparative statics*. It is apparent from this equation that the inverse of the *bordered Hessian* or

$$H^{-1} = \begin{bmatrix} L_{xx} & g_x^T \\ g_x & 0 \end{bmatrix}$$

is the key to comparative static results in the "classical approach". In this approach second order necessary conditions provide *local* information and concavity of f and the g^i's provide *global* information about this matrix.

□ *Problem* 5.10. Suppose that the unconstained optimization problem, $\max_x f(x, a)$, has a unique solution $x^*(a)$ where $x, a \in \mathbb{R}$ and $\partial^2 f(x, a)/\partial x^2 < 0$.

1. Show that $\text{Sign}(dx^*(a)/da) = \text{Sign}(\partial^2 f(x, a)/\partial x \partial a)$ where $\partial^2 f(x, a)/\partial x \partial a$ is evaluated at $x = x^*(a)$.

2. Interpret this result geometrically by plotting $\partial f(x, a)/\partial x$ for a given value of a in a graph with x plotted on the horizontal axis.

5.3.2 The Envelope Theorem

An important alternative approach to comparative statics is provided by the envelope theorem. Consider a controlled experiment in which only the ith exogenous variable will be changed. The envelope theorem concerns the following question. How does

$$\begin{aligned} F(a) &\equiv f(x(a), a) \\ &= f(x(a), a) + \lambda(a) g(x(a), a) \end{aligned} \tag{5.13}$$

change with a_i? Put more precisely, what is the (partial) derivative of $F(a)$ with respect to a_i? Consider the right hand side of this expression and notice that as a_i changes there are two types of effects upon the value of $f(x(a), a)$. The first corresponds to the (partial) derivative of $f(x, a)$ with respect to a_i evaluated at $x = x(a)$. This is the direct effect and is denoted $\partial f(x(a), a)/\partial a_i$. The second, or indirect effect, corresponds to the effect of the change in a_i upon the components of x, $\partial x_j(a)/\partial a_i$, and the effect of these changes in the components of x upon $f(\cdot)$, $\partial f(\cdot)/\partial x_j$. Using the chain rule, the total effect is the sum of these two components:

$$\frac{\partial F(a)}{\partial a_i} \equiv \frac{\partial f(a)}{\partial a_i} + \sum_{j=1}^{n} \frac{\partial f(x, a)}{\partial x_j} \frac{\partial x_j(a)}{\partial a_i}$$

The envelope theorem says, quite simply, that the second — and complicated — term in this expression is zero:

Theorem 30 (Envelope Theorem). Suppose

$$\begin{aligned} x(a) \equiv \max_x & f(x, a) \\ \text{s.t. } & g(x, a) \geq 0 \end{aligned}$$

and

$$\begin{aligned} F(a) &= f(x(a), a) \\ L(a) &= f(x(a), a) + \lambda(a) g(x(a), a) \end{aligned}$$

where $x(a)$ and $\lambda(a)$ are functions satisfying the Kuhn-Tucker conditions (Equation 5.4 and Equation 5.7 on page 61). Provided that both $F(\cdot)$ and $L(\cdot)$ are differentiable

$$\frac{\partial F(a)}{\partial a_i} = \frac{\partial L(x, \lambda, a)}{\partial a_i}$$

where

$$\frac{\partial L(x, \lambda, a)}{\partial a_i} = \frac{\partial f(x, a)}{\partial a_i} + \lambda \frac{\partial g(x, a)}{\partial a_i}$$

and all partial derivatives are evaluated at $x = x(a)$ and $\lambda = \lambda(a)$.

The basis of this surprising fact and the reason it is called the envelope theorem is illustrated in Figure 5.4. In this example there are no constraints and a has a single component:

$$\begin{aligned} f(x,a) &= \frac{[a-(x-a)^2]}{2} \\ x^*(a) &= \arg\max_x \frac{[a-(x-a)^2]}{2} \\ &= a \\ F(a) &= \max_x \frac{[a-(x-a)^2]}{2} \\ &= f(x^*(a),a) \\ &= \frac{a}{2} \end{aligned}$$

In Figure 5.4 the curve $F(a)$ describes the value of the objective function as a changes and as x is continually changed to remain optimal for the new values of a. The slope of $F(a)$ at a point corresponds, of course, to the total derivative $dF(a))/da$. A series of other curves are plotted in which x is held constant at the values which would be optimal for various specific values of a, e.g., $f(7,a)$ corresponds to holding x constant at $x = 7$ — the value of x which is optimal when $a = 7$ — and then varying a. The slope of one of these fixed x curves at a particular point is the partial derivative $\partial f(x,a)/\partial(a)$.

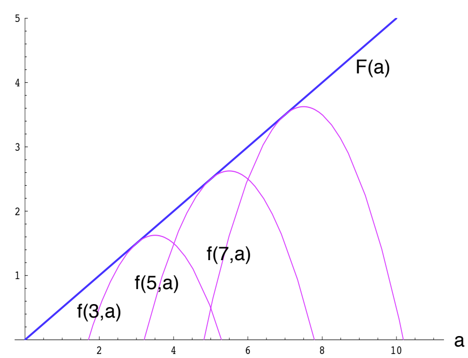

Figure 5.4: The Envelope Theorem without Constraints

Note that the graph of $F(a)$ is the "envelope" of the fixed x curves. The reason for this is simple. None of the fixed x curves can every lie above $F(a)$ which, after all, is the *maximized* value of $f(x,a)$ over all possible choices of x. The fixed x curves can touch $F(a)$, however. Consider the point at which $a = 7$ for example. Since $x = 7$ is optimal for this value of a, it follows that $F(7) = f(7,7)$. The fixed x curve $f(7,a)$ thus touches $F(a)$ at $a = 7$. A similar story holds for the other fixed x curves each of which contributes, in this case, a single point to the envelope $F(a)$.

Now if two curves touch at a point but do not cross they are tangent to one another. If, moreover, both curves are differentiable, then their slopes are well defined at all points and their slopes must be equal at points of tangency. This is the envelope theorem — the slope of a fixed x curve must be the same as the slope of the envelope curve at the point of tangency.

A slightly more complex illustration of the envelope theorem is provided by the following problem:

$$\begin{aligned} &\arg\max_x x^2 \\ &\qquad \text{s.t. } x \le a \end{aligned}$$

The Lagrangian function for this problem is

$$L(x,\lambda,a) \equiv x^2 + \lambda(a-x)$$

and the solution is given by

$$\begin{aligned} x^*(a) &= a \\ \lambda^*(a) &= 2a \\ F(a) &\equiv L(x^*(a), \lambda^*(a), a) \\ &= a^2 \end{aligned}$$

This problem is illustrated in Figure 5.5. Note that the "fixed x" curves are obtained by holding both x and λ constant at values which would be optimal for some a and then letting a vary. Note again that $F(a)$ is the envelope of these curves and that, for example, the tangency of $F(a)$ with $L(3,6,a)$ at $a = 3$ means that

$$\frac{dF(3)}{da} = \frac{\partial L\left(x^*(3), \lambda^*(3), 3\right)}{\partial a}$$

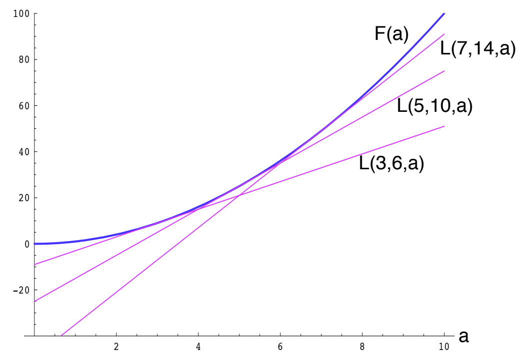

Figure 5.5: The Envelope Theorem with Constraints

□ *Problem* 5.11. Consider the problem

$$\begin{aligned} &\max_{x \geq 0} \; \min\{x_1, x_2\} \\ &\text{s.t. } p_1 x_1 + p_2 x_2 \leq m \end{aligned}$$

with parameters $p_1, p_2, m > 0$.

1. Show that the solution to this problem is given by

$$\begin{aligned} x_i^*(p_1, p_2, m) &= \frac{m}{p_1 + p_2}, \quad i = 1, 2 \\ F(p_1, p_2, m) &= \frac{m}{p_1 + p_2} \end{aligned}$$

2. Is the objective function differentiable? Is $F(p_1, p_2, m)$ differentiable?

□ *Problem* 5.12. Consider the problem

$$\begin{aligned} &\max_{x \geq 0} x_1 + x_2 \\ &\text{s.t. } p_1 x_1 + p_2 x_2 \leq m \end{aligned}$$

with parameters $p_1, p_2, m > 0$.

1. Show that a solution to this problem is given by

$$x_1^* = \begin{cases} 0 & \text{if } p_1 > p_2 \\ m/p_1 & \text{if } p_1 \leq p_2 \end{cases}$$

$$x_2^* = \begin{cases} m/p_2 & \text{if } p_1 > p_2 \\ 0 & \text{if } p_1 \leq p_2 \end{cases}$$

2. Is the objective function differentiable? Is $F(p_1, p_2, m)$ differentiable?

□ *Problem* 5.13. Solve Problem 5.11 using *Mathematica* for the case in which $p_1 = 2$, $p_2 = 3$ and $m = 10$.

5.3.3 The Envelope Approach

The connection of the Envelope Theorem to comparative statics depends on a special characteristic of problems frequently encountered in microeconomics. The endogenous variables in such problems are the quantities chosen of various commodities and the exogenous parameters are the market prices of these same commodities. The objective function, moreover, commonly depends upon the market value of the quantities chosen — the dot product of prices and quantities. Such problems have the special form:

$$\begin{aligned} &\max_x a \cdot x \\ &\text{s.t. } g(x) \geq 0 \end{aligned} \tag{5.14}$$

Since the Lagrangian for Equation 5.14 is

$$L(x, \lambda, a) = a \cdot x + \lambda g(x)$$

the Envelope Theorem implies that

$$\begin{aligned} \frac{\partial F(a)}{\partial a_i} &= \frac{\partial L(x, \lambda, a)}{\partial a_i} \\ &= x_i(a) \end{aligned}$$

and

$$\frac{\partial^2 F(a)}{\partial a_i \partial a_j} = \frac{\partial x_i(a)}{\partial a_j}$$

Here the comparative static effects, $\partial x_i(a)/\partial a_j$, are obtained as the second partial derivatives of $F(a)$.

In this approach *global* information about the concavity or convexity of $F(a)$ provides *global* information about the comparative static effects.