Chapter 2

Constrained Optimization

2.1 Introduction

This chapter introduces and explains the fundamental mathematical tools that we will need to study consumer theory. It is a self contained summary of multivariable calculus and constrained optimization. It assumes familiarity with calculus of a single variable.

Multivariable calculus is a pre requisite to understanding constrained optimization which is the fundamental technique that economists use to analyze economic problems. Consider the economic problem faced by consumers, the subject of Part I. Consumers are rational in the sense that they use their limited resources to obtain the maximum amount of happiness that they can. They must choose the combination of consumption that grants them the highest possible level of utility. Choice and trade-offs are at the root of any economic problem. If consumers were to get utility from only one good, the solution to the economic problem would be trivial: consume as much as possible. In order to have a meaningful economic problem, we need to provide consumers with a choice. In this case the solution to the economic problem is not as obvious. Constraints illustrate the fact that resources are limited and the Calculus is the tool that we use to implement the rationality of consumers. Constrained optimization is, therefore, the way in which we study the interaction between scarcity and rational choice.

We will discuss two solution methods to a constrained optimization problem. First, we will study the **substitution method**, which essentially eliminates the constraint and one variable; then, we will proceed to study the **method of Lagrange multipliers** which introduces an extra variable and treats the constraint explicitly. We will also explain why these two methods are equivalent. That is, why they give the same solution to well posed economic problems. However, we will highlight the benefit of each solution method, in terms of the economic intuition that we derive from each one.

We will use techniques from the Calculus to attack the problem of **comparative statics** next. Comparative statics concerns the effect that a change in parameters has on the solution to our problem.

2.2 Partial Derivatives

Suppose that we have a function of two variables. Call it $f(x, y)$. We want to know how rapidly the function changes when the variables change. The problem is that if both variables change at the same time, it may be difficult to isolate the effect of, say, the change in x. **Partial derivatives** allow us to quantify the effect on a function of the change in one variable when all others held constant.

Suppose we take the function $f(x, y)$, but we evaluate it at a fixed value of y, say $\bar{y}$. Now we have a function of only one variable, since now the value of y is fixed, and all the differentiation rules from single variable calculus apply. Our function is now equal to:

$$f(x, y = \bar{y}).$$

Suppose that I want to know how rapidly the function is changing in the x direction. Then, I take the derivative with respect to x of $f(x, y = \bar{y})$:

$$f_x = \frac{\partial f(x, y)}{\partial x} = \frac{df(x, y = \bar{y})}{dx}.$$

Notice that what we are doing is holding the value of y constant and taking the derivative with respect to x. The derivative will, in general, be a function of both x and y. This means that the magnitude of f_x depends on the value of y at which we evaluate f_x.

Example 1 *Suppose the* $f(x, y) = x^{\alpha} y^{1-\alpha}$. *Then,*

$$f_x = \frac{\partial f}{\partial x} = \alpha x^{\alpha-1} y^{1-\alpha}$$

and

$$f_y = \frac{\partial f}{\partial y} = (1 - \alpha)\, x^{\alpha} y^{-\alpha}.$$

Notice how we hold constant one of the variable and then we look for the magnitude of change in the direction of the other variable, treating the function as if it were a single variable function.

2.3 Total Derivatives

Suppose, now, that both x and y change simultaneously.. We want to find the effect of this change on f. We will compute the **total derivative** of f by decomposing the total change in f as the sum of two partial changes. First, hold x fixed and compute the change in the direction of y. This is f_y. Multiply f_y by the magnitude of the change, say, dy. Then, $f_y dy$ is the magnitude of the change in f due to the change in y holding x constant. To find the total change, add to this quantity the effect of the change in x holding y constant: $f_x dx$. The total differential is equal to:

$$df = f_x dx + f_y dy.$$

Suppose that we divide both sides by dx :

$$\frac{df}{dx} = f_x + f_y \frac{dy}{dx}.$$

This expression tells us that the total change in f due to a change in x is the sum of a **direct effect** and an **indirect effect**. The direct effect, f_x, holds y fixed. The indirect effect is the extra change that arises from the effect that x may have on y, $\frac{dy}{dx}$, and the subsequent effect that y has on f, holding x constant, f_x. Suppose that we take y to be constant as we do in partial differentiation. Then, $\frac{dy}{dx} = 0$ and the indirect effect vanishes. The change in f is simply given by the direct effect which is the partial derivative with respect to x :

$$\frac{df}{dx} = f_x$$

the total effect is simply the partial effect.

2.4 The Substitution Method

The standard constrained optimization problem is given by:

$$[P]: \max_{\{x,y\}} \quad f(x,y)$$
$$s.t. \qquad g(x,y) = m$$

Notice that there are three variables in this problem. We wish to select the **control variables** $\{x, y\}$ to make the objective function, $f(x,y)$, as large as possible, as long as when we plug these values into the constraint, $g(x,y)$, we obtain exactly m. The third variable, m, is a **parameter**, which we may take as a given of the problem.

The **substitution method** is very intuitive. First we solve the constraint for one of the control variables, say y, as a function of the other control variable, x, and the parameter m:

$$y = h(x,m).$$

Next we substitute y out of the problem to obtain the following optimization problem, which we label $[S]$:

$$[S]: \max_{\{x\}} \quad f(x, h(x,m))$$

The substitution method allows us to convert a problem that we are not quite sure how to solve into a single variable calculus problem that we can solve by taking a derivative. The first order condition of this problem is given by the following expression:

$$[x]: \quad f_x + f_y \tfrac{dh}{dx} \quad .$$

Notice that we use the chain rule to obtain the first order condition. We need it because the constraint does not allow us to select x and y independently of each other; any optimal choice must satisfy the constraint. Therefore, when we vary x a little bit, we need to consider the effect that varying x has on our choice of y. In other words, because the solution must satisfy the constraint the effect that changing x has on the feasible value of y (this happens through $\frac{dh}{dm}$) must affect our choice of x.

There are several points that deserve special mention. First, two letters appear in the first order condition: x and m. But, remember that m is a parameter so it is given to us and we consider it a number. Therefore, the solution to the problem is given by the value of x that makes the first order condition exactly equal to 0. It turns out that since we do not have a value for the parameter m, the maximizing value of x will be a function of m. Also, we will place a star on the maximizing value of x, x^*, to indicate that x^* is a number. Therefore, the maximizing value of x is written as $x^*(m)$ and it solves the following equation:

$$f_x + f_y \frac{dh}{dx} = 0.$$

To obtain the maximizing value of y, we plug x^* into the implicit function we obtained above:

$$y^*(m) = h(x^*(m), m).$$

Therefore, the solution to $[P]$ when we use $[S]$ is given by two equations:

$$\begin{aligned} f_x + f_y \frac{dh}{dx} &= 0 \\ y^*(m) &= h(x^*(m), m). \end{aligned}$$

The first condition determines the maximizing value of x and the second utilizes the constraint to obtain the maximizing value of y. It is important to note that we use the constraint twice. First we use it in the maximization to replace the value of one variable. Second, we use it to recover the optimal value of the variable we substituted out.

Consider the following simple example.

Example 2 *Suppose we want to solve the following maximization problem:*

$$\begin{aligned} \max_{x,y} \quad & \log(x) + \log(y) \\ s.t. \quad & x + y = m \end{aligned}$$

Then,

$$\begin{aligned} f(x,y) &= \log(x) + \log(y) \\ g(x,y) &= x + y \end{aligned}$$

Solving for $h(x,m)$, we obtain

$$y = m - x = h(x,m),$$

and substituting for y we obtain the following equivalent maximization problem:

$$\max_x \quad \log(x) + \log(m - x)$$

The first order condition for this problem is

$$[x]: \quad \tfrac{1}{x^*} + \tfrac{1}{m-x}(-1) \quad = \quad 0 \ .$$

and solving for x^* we obtain

$$x^* = \frac{m}{2}.$$

Then, we use $y = h(x, m) = m - x$ to obtain the optimal value of y :

$$y^* = \frac{m}{2}.$$

Notice that as we stated previously, the solution expresses the choice variables, in this case x and y, as functions of the parameters, in this case m.

Alternatively we could substitute in for x after taking the first order conditions. We show the equivalence of both procedures in the following example.

Example 3 *Consider the same maximization as in the previous example. In the text we derived the two conditions that determine the optimum values of x and y, which we reproduce below for convenience:*

$$y^*(m) = h(x^*(m), m)$$
$$f_x + f_y \frac{dh}{dx} = 0.$$

The first condition is a simple manipulation of the constraint, evaluated at the optimal value. The second is the first order condition with respect to x, also evaluated at the optimum. In our simple maximization problem the first condition is given by

$$y^* = m - x^*$$

and the second is given by

$$\frac{1}{x^*} + \frac{1}{y^*}(-1) = 0.$$

If we substitute for y^* in the second condition, the equations determining the optimum values of x and y are given by:

$$\frac{1}{x^*} + \frac{1}{m - x^*}(-1) = 0$$
$$y^* = m - x^*$$

which is precisely the first order condition that with respect to x of the previous problem.

2.5 The Method of Lagrange Multipliers

Instead of replacing the constraint into the objective function, the **method of Lagrange multipliers** introduces one more variable, λ, into the problem. This variable is known as the **Lagrange multiplier** and has an important economic interpretation which we will get to later. The method of Lagrange relies on maximizing an associated function, called the Lagrangian. We form the Lagrangian by adding λ times the constraint to the objective function and maximizing over the control variables and also the Lagrange multiplier:

$$[L]: \max_{\{x,y,\lambda\}} \quad f(x,y) + \lambda(m - g(x,y)) \ .$$

In order to maximize the Lagrangian, we take partial derivatives with respect to the three control variables:

$$\begin{aligned} [x]: &\quad f_x - \lambda g_x \\ [y]: &\quad f_y - \lambda g_y \\ [\lambda]: &\quad m - g(x,y) \end{aligned}$$

and to obtain the maximizing values of the controls, x^*, y^*, λ^*, we set these equations equal to zero and solve them simultaneously.

First we note that this method treats the constraint explicitly, therefore, there is no chain rule effect in the first order conditions. In the process of taking first order conditions, we treat all variables except the control variable at hand as constant. Thus, we take only partial derivatives and not total derivatives. In this case, the solution satisfies the constraint because the constraint is one of the conditions that must be explicitly satisfied.

Example 4 *Consider the same maximization problem that we solved via the substitution method in the previous section. That ism suppose we want to solve*

$$\begin{array}{cc} \max_{x,y} & \log(x) + \log(y) \\ s.t. & x + y = m \end{array}$$

using Lagrange multipliers. First we form the Lagrangian

$$[L]: \max_{\{x,y,\lambda\}} \log(x) + \log(y) + \lambda(m - x - y) .$$

Now we find the first order conditions by taking derivatives with respect to x, y, *and* λ *and setting them equal to* 0 :

$$\begin{array}{ll} [x]: & \frac{1}{x^*} - \lambda^* = 0 \\ [y]: & \frac{1}{y^*} - \lambda^* = 0 \\ [\lambda]: & m - x^* - y^* = 0. \end{array}$$

Now we obtain the optimality condition by eliminating λ^* *from* $[x]$ *and* $[y]$:

$$x^* = y^*.$$

This give us the relationship between x *and* y *when they are optimally chosen.* $[\lambda]$ *is the feasibility condition because it pins down the level of* x *and* y *consistent with the available resources:*

$$x^* + y^* = m.$$

We have two equations and two unknowns which we can solve for x^* *and* y^*. *The solution is given by:*

$$x^* = y^* = \frac{m}{2},$$

precisely what we obtained with the substitution method. Recall, however, that λ *is also a choice variable and, consequently, we must solve for it too. We can obtain it either from* $[x]$ *or* $[y]$:

$$\lambda^* = \frac{2}{m}.$$

Therefore, notice that both methods give us the same solution, but Lagrange's method give us more information! In addition to the optimal values of x and y, it also tells us the value of λ. We will explore the reasons for the equivalence of the two problems in the following section. Then, we will provide the economic interpretation of the Lagrange multiplier.

2.6 Equivalence of [S] and [L]

Both methods are equivalent if they yield the same solution to the constrained optimization problem. That is we should get the same optimal values of x and y. This will occur if both methods have the same first order conditions or if we can obtain the first order conditions of one method from the first order conditions of the other. Certain conditions must be satisfied in order for the solution to the problem to exist. We will leave these restrictions unstated but we will provide an example where the solution method fails after the simple proof. The type of problem that we are interested in solving will not be problematic so you can use either method with confidence.

Now, on to the proof. We want to show that [x] and [y] of [L] are equivalent to [x] of [S]. We will begin by showing that we can get the necessary conditions of [S] from those of [L]. Therefore we need to eliminate λ and consolidate the remaining conditions into one equation that allows us to recover the value of the control variable from [S]. Thus, we look for a value of λ that leads both sets of first order conditions to be equivalent. In both cases, the constraint provides the value of the other control.

First, we set the first order conditions of [L] equal to zero. Let us solve [y] for λ^* and obtain:

$$\lambda^* = \frac{f_y}{g_y}.$$

If we plug in this value of λ^* into [y] for [L] the first order condition holds trivially:

$$f_y - \frac{f_y}{g_y} g_y = 0$$

because we used that same equation to find λ^* so it must hold. If we replace the value for λ^* into [x] of [L] we obtain:

$$f_x - \frac{f_y}{g_y} g_x$$

Recall that all partials are functions of x^* and y^* only.

Now, we look at [S]. The first order condition of this problem is given by:

$$f_x + f_y \frac{dh}{dx} = 0,$$

where

$$y = h(x, m).$$

We can use implicit differentiation on the constraint to obtain $\frac{dy}{dx}$. We know that:

$$g_x dx + g_y dy = 0$$

and solving for $\frac{dy}{dx}$ we obtain:

$$\frac{dy}{dx} = -\frac{g_x}{g_y} = \frac{dh}{dx}$$

where the last equality follows because we obtained the implicit function $h(x, m)$ by solving the constraint for y. Notice that there is an implicit assumption here: $g_y \neq 0$ must hold everywhere (for every possible x, y combination) because otherwise we would be dividing by 0 and the entire procedure would break down.

Since $\frac{dy}{dx} = -\frac{g_x}{g_y}$, we can rewrite the first order condition we obtained from [S] as:

$$f_x + f_y\left(-\frac{g_x}{g_y}\right) = 0,$$

which we can re-write as:

$$f_x - \left(\frac{f_y}{g_y}\right) g_x = f_x - \lambda^* g_x = 0$$

so the first order condition [x] of [L] holds too. The last piece of information is provided by the constraint which in [S] we substituted into the first order condition. We need it to recover the optimal value of y, y^*. In [L], we consider it explicitly and use it, also to obtain the optimal value of y. Thus, we have shown that we can get [S] from [L]. We can show that we can get [L] from [S] by working backwards. This establishes the equivalence of [S] and [L].

2.7 The Economic Interpretation of λ

If [S] and [L] are equivalent, then why bother with [L]? The answer to this question is that λ provides important economic information. Remember that we will be interested in obtaining information about behavior, so that we are not done once we have obtained the solution. In addition, we want to know how the solution varies when we change the parameters of the problem; that is we wish to compute **comparative statics** properties. λ is the answer to one such question.

λ is often referred to as the shadow value of the constraint (in economic applications it is called the marginal utility of income). It tells us by how much the maximum value of the objective function changes when we relax the constraint by one unit:

$$\frac{df^*}{dm} = \lambda.$$

We know that the solution to $[L]$ can be written as $x^*(m)$ and $y^*(m)$, so that the maximum value of the objective function is obtained by plugging the maximizers into it:

$$f^*(m) = f(x^*(m), y^*(m)).$$

If we take the derivative of the objective function with respect to m, we obtain:

$$\frac{df^*}{dm} = f_x^* \frac{dx^*}{dm} + f_y^* \frac{dy^*}{dm},$$

but if we recall that $[x]$ and $[y]$ of $[L]$ are given by

$$\begin{aligned} [x]: &\quad f_x^* - \lambda^* g_x^* = 0 \\ [y]: &\quad f_y^* - \lambda^* g_y^* = 0 \end{aligned}$$

we can re-write the derivative as:

$$\frac{df^*}{dm} = \lambda^* \left(g_x^* \frac{dx^*}{dm} + g_y^* \frac{dy^*}{dm} \right),$$

by solving the first order conditions for f_x^* and f_y^* and, then, substituting them out in the expression for $\frac{df^*}{dm}$ above.

We also know that the constraint must hold with equality:

$$g(x^*(m), y^*(m)) = m.$$

Therefore, if we take the derivative of the constraint with respect to m, we obtain

$$g_x^* \frac{dx^*}{dm} + g_y^* \frac{dy^*}{dm} = 1,$$

which we can use to substitute and obtain the desired result:

$$\frac{df^*}{dm} = \lambda^*.$$

This result is an example of comparative statics. It tells us how the solution to [P] varies as we vary the parameters of the problem. Throughout the course we will be interested in studying the properties of a two versions of [P] : the problems of utility maximization and expenditure minimization.

2.8 The Envelope Theorem

2.8.1 Unconstrained Optimization

Suppose that we solve the following maximization problem:

$$\max_x f(x; a),$$

where x is a choice variable and a is a parameter. The first order condition of this problem is:

$$f_x(x^*; a) = 0.$$

The optimal choice of x is a function of $a : x^*(a)$. Then the maximum value of f is given by:

$$f(x^*(a), a),$$

which is only a function of a.

Now we want to see the effect that a change in a has on $f(x^*(a), a)$. Therefore, we compute the total derivative. We not only want the direct effect of a, but also the effect that a has on f, through x^* :

$$\frac{df}{da} = f_a + f_x \frac{dx}{da}.$$

Notice, though, that the first order condition of the maximization states that $f_x = 0$. Therefore,

$$\frac{df}{da} = f_a.$$

That is, the total effect is equal to the partial effect! This happens because when we vary a we adjust the choice of x^* optimally. This requires that $f_x = 0$ at the new choice. Consequently, the envelope theorem tells us that when we look for the derivative of a maximum value function with respect to a parameter, we do not need to worry about the indirect effects of this parameter, but we can focus only the direct effect. The reason: choice variables will adapt to their optimal values, so that first order conditions will be satisfied at every point of the maximum value function.

2.8.2 Constrained Optimization

Suppose, now, that we have a function of two variables, subject to a constraint. Let us verify the envelope theorem in this case. By establishing this result we will understand the significance of the Lagrange multiplier from another very useful perspective. Let's consider the standard maximization that we have been studying throughout this chapter:

$$\begin{array}{ll} \max\limits_{x,y} & f(x,y) \\ s.t. & g(x,y) = m \end{array}$$

We wish to consider the effect that varying m has on the maximum attainable value of $f(x,y)$.

If we solve the maximization problem using Lagrange multipliers we form the Lagrangian

$$[L]: \max_{\{x,y,\lambda\}} f(x,y) + \lambda(m - g(x,y)) .$$

Then we find the first order conditions by taking derivatives with respect to x, y, and λ and setting them equal to 0:

$$\begin{array}{ll} [x]: & f_x(x^*(m), y^*(m)) - \lambda^*(m) g_x(x^*(m), y^*(m)) = 0 \\ [y]: & f_y(x^*(m), y^*(m)) - \lambda^*(m) g_y(x^*(m), y^*(m)) = 0 \\ [\lambda]: & m - x^*(m) - y^*(m) = 0. \end{array}$$

As the above first order conditions make clear, the solution is the set of choice variables, x, y, and λ, as functions of the parameters, which in this case is m. These are the values of the choice variables that satisfy the first order conditions at equality.

If we evaluate the objective function at the optimal values we obtain the **maximum value function:**

$$f^*(m) = f(x^*(m), y^*(m))$$

$$
\begin{aligned}
&= L(x^*(m), y^*(m), \lambda^*(m)) \\
&= L^*(m).
\end{aligned}
$$

where the second equality follows because **at the optimum** the constraint is satisfied with equality, so we are adding zero to the maximized value of $f(x,y)$. We wish to determine the effect that varying m has on $f^*(m)$. In other words, we want to find an expression for $\frac{df^*(m)}{dm}$. By the second equality above, we know that

$$\frac{df^*(m)}{dm} = \frac{dL^*(m)}{dm}.$$

The envelope theorem tells us that

$$\frac{dL(x^*(m), y^*(m))}{dm} = \frac{\partial L(x^*(m), y^*(m), \lambda^*(m))}{\partial m}.$$

In other words, the total effect that a change in m has on the optimized value of the Lagrangian is equal to the partial effect that m has on the optimal value of the Lagrangean. Since at the optimum the maximum value of the Lagrangian is equal to the maximum value of f, the envelope theorem, applied to the Lagrangian establishes that:

$$\frac{df^*(m)}{dm} = \frac{dL^*(m)}{dm} = \lambda^*(m).$$

Let's explore why the envelope theorem holds. If we take the total derivative of $L(x^*(m), y^*(m), \lambda^*(m))$ with respect to m, we obtain must consider the direct and the indirect effects. The total derivative would, then, be the sum of both. The total derivative is given by:

$$\frac{dL(x^*(m), y^*(m), \lambda^*(m))}{dm} = \frac{\partial L^*}{\partial x}\frac{dx^*(m)}{dm} + \frac{\partial L^*}{\partial y}\frac{dy^*(m)}{dm} + \frac{\partial L^*}{\partial \lambda}\frac{d\lambda^*(m)}{dm} + \frac{\partial L^*}{\partial m}.$$

where the first three terms are the indirect effect –they quantify the effect that changing m has on L^* through x, y, and λ – while the last term is the direct effect that m has on L^*. However, the first three terms must equal zero because they precisely the three first order conditions evaluated at the maximum:

$$
\begin{aligned}
\frac{\partial L^*}{\partial x} &= f_x^* - \lambda^* g_x^* = 0 \\
\frac{\partial L^*}{\partial y} &= f_y^* - \lambda^* g_y^* = 0 \\
\frac{\partial L^*}{\partial \lambda} &= m - g(x^*, y^*) = 0
\end{aligned}
$$

Also, note that the direct effect is:

$$\frac{\partial L^*}{\partial m} = \lambda^* (m) .$$

Putting all these results together, the **envelope theorem for constrained optima** tells us that:

$$\frac{dL^* (m)}{dm} = \frac{\partial L^* (m)}{\partial m}.$$

But, since at the maximum the optimal value of the Lagrangian equal the optimized value of the objective function:

$$\frac{df^* (m)}{dm} = \frac{\partial L^* (m)}{\partial m}.$$

Since the direct effect is simply the multiplier we obtain the envelope theorem in its usual form:

$$\frac{df^* (m)}{dm} = \lambda^* (m)$$

Economically speaking the Lagrange multiplier can be interpreted as the **shadow price** associated with the constraint. It tells us how much we would be willing to pay to have one more unit of m. Alternatively, it tells us the cost of violating the constraint infinitesimally. It tells us the cost of using slightly more m or, equivalently, it tells us the cost of using all but a tiny little bit of m.

Finally, one remark on the mathematical argument we have just used to establish our result. We did not use any math that we did not use in the unconstrained case. The key insight is to realize that since at the optimum the constraint must hold with equality, we can look at the total derivative of the optimized Lagrangian instead of the optimized value of the objective function. This key step allows us to transform a constrained problem into an unconstrained one! We learned how to analyze this problem in the previous section.

Example 5 *Let's continue with the problem we have been solving throughout this chapter.*

2.9 Second Order Conditions

To be completed.

2.10 The Kuhn-Tucker Theorem

To be completed.

2.11 Exercises

Exercise 2.11.1 *In the text we solved*

$$\begin{array}{ll} \max\limits_{x,y} & \log(x) + \log(y) \\ s.t. & x + y = m \end{array}$$

via the substitution method and via Lagrange's method. Follow the proof of the equivalence of both methods to establish equivalence in the context of this problem. That is, show that you can obtain the first order conditions of $[S]$ *from the first order conditions of* $[L]$ *and vice-versa.*