Economics 5113
Introduction to Mathematical Economics
Winter 1999

Lecture 12
Constrained Optimization

I. Introduction

A. Due largely to the importance of maximization of utility subject to a budget constraint, constrained optimization has achieved a prominent place in economic theory. Other problems for which it is the appropriate framework include:

1. Maximizing the efficiency of the tax system subject to a revenue constraint.
2. Minimizing the cost of a compensation package subject to a constraint on employee utility.
3. Maximizing the utility of a “representative consumer” subject to a technological constraint describing the trade-off between economic output and environmental quality.

B. We will develop the Lagrangean method of analysis.

1. The general theory will be presented and, to some extent, explained.

C. My own view of this material is:

1. The Lagrangean methodology is a general, theoretical description of the state of affairs at a solution to an optimization problem.
   a. In economic theory it is particularly important that the Lagrangean multipliers can be interpreted as shadow prices expressing how the value of the problem would change as the constraint is tightened or
loosened.

b. The advantage of generality is that one has a single machine capable of handling a wide variety of problems. This is particularly convenient for computers.

2. When you are dealing with any particular problem the mechanical application of Lagrangean methods is cumbersome and unreliable.

a. In comparison with using the constraint to express one of the variables in terms of the others, substituting, and solving the derived unconstrained optimization problem, the Lagrangean technique involves more variables and more equations, which must be solved by substitution anyway, so that the chances of human error are greater.

b. The “advantage” of the Lagrangean method is its universality and mechanical quality. That is, you don’t have to think about what type of problem you’re trying to solve, and you don’t have to think about how to solve it. In fact, though, the problem is rarely simplified or clarified by being put in Lagrangean form.

II. The Lagrangean Theorem for a Single Constraint

A. Statement and proof.

Theorem: let \( f : \mathbb{R}^n \to \mathbb{R} \) and \( g : \mathbb{R}^n \to \mathbb{R} \) be \( C^1 \) functions. If \( x^* \) solves the problem

\[
(*) \quad \max f(x) \quad \text{subject to} \quad g(x) = 0
\]

and \( \frac{\partial g}{\partial x_i}(x^*) \neq 0 \) for some \( i \), then there is a \( \lambda^* \in \mathbb{R} \) such that, with \( L(x, \lambda) = f(x) + \lambda g(x) \),

\[
\frac{\partial L}{\partial x_i}(x^*, \lambda^*) = 0 \quad (i = 1, \ldots, n) \quad \text{and} \quad \frac{\partial L}{\partial \lambda}(x^*, \lambda^*) = 0.
\]

Proof: We have \( \frac{\partial g}{\partial x_i}(x^*) \neq 0 \) for some \( i \); renumbering the coordinates, we may assume that \( \frac{\partial g}{\partial x_n}(x^*) \neq 0 \). We write \( x^* = (y^*, z^*) \) where \( y^* \in \mathbb{R}^{n-1} \) and \( z^* \in \mathbb{R} \). The implicit
function theorem now states that there is $U \subseteq \mathbb{R}^{n-1}$, a neighborhood of $y^*$, and a $C^1$ function $h : U \to \mathbb{R}$ such that:

(a) $h(y^*) = z^*$;

(b) $g(y, h(y)) = 0$ for all $y \in U$.

Since $x^*$ is a solution to (*), $y^*$ is necessarily a solution to

\[(**)
\max_{y \in U} f(y, h(y)).\]

The first order conditions for this problem are that

\[\frac{\partial f}{\partial y_j}(y^*, z^*) + \frac{\partial f}{\partial z}(y^*, z^*) \cdot \frac{\partial h}{\partial y_j}(y^*) = 0 \quad (j = 1, \ldots, n - 1),\]

and the implicit function theorem states that

\[\frac{\partial h}{\partial y_j}(y^*) = -\frac{\partial g}{\partial y_j}(y^*, h(y^*)) \quad (j = 1, \ldots, n - 1).\]

Obviously we must set

\[\lambda^* = -\frac{\partial f}{\partial z}(y^*, z^*)\]

to have $\frac{\partial \mathcal{L}}{\partial x_i} = 0$ for $i = 1, \ldots, n - 1$. For $i = n$ we have

\[\frac{\partial \mathcal{L}}{\partial x_n} = \frac{\partial f}{\partial x_n} - \frac{\partial f}{\partial n} \cdot \frac{\partial g}{\partial x_n} = 0\]

since $z = x_n$. Of course $\frac{\partial \mathcal{L}}{\partial x_n}(x^*, \lambda^*) = g(x^*) = 0$ by construction.

\[\square\]

B. Geometric Interpretation

1. The vector

\[\frac{\partial f}{\partial x}(x^*) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x^*) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x^*) \end{bmatrix}\]

is called the gradient of $f$ at $x^*$.

a. The normalized gradient is the unit vector $\frac{\frac{\partial f}{\partial x}(x^*)}{\|\frac{\partial f}{\partial x}(x^*)\|}$ in the direction in which $f$ increases most rapidly.
b. The gradient is consequently the normalized gradient multiplied by
the rate at which \( f \) increases as one moves, at unit speed, in the
direction of the normalized gradient.

c. Geometrically, the Lagrangean theorem asserts that, so long as the
gradient of \( g \) does not vanish, the gradient of \( f \) is a scalar multiple
of the gradient of \( g \).

2. Observe that if \( v \in \mathbb{R}^n \) is a vector, which we imagine to be “located” at
\( x^* \), that is tangent to the surface given by the equation \( g(x) = 0 \), then
\[
0 = \sum_{i=1}^{n} \frac{\partial g}{\partial x_i}(x^*)v_i = \frac{\partial g}{\partial x}(x^*) \cdot v,
\]
so that \( v \) and \( \frac{\partial g}{\partial x}(x^*) \) are perpendicular.

a. Thus we can think of \( \frac{\partial g}{\partial x}(x^*) \) as a vector that is perpendicular to
the surface \( M = \{ x \in \mathbb{R}^n : g(x) = 0 \} \). (To be ultra-precise we
would say that the gradient of \( g \) is perpendicular to the “tangent
plane” of the surface \( M \).

b. We can now see the Lagrangean theorem as asserting that the gra-
dient of \( f \) is perpendicular to \( M \).

c. There is a very important improvement here. The mathematical
role of \( g \) is entirely a matter of specifying \( M \) as the set of solutions
of an equation, so it is good to have achieved a statement of the the-
orem that refers only to the essential element of the picture, namely
\( M \), without mentioning inessential information, namely the partic-
ular function \( g \), which is not better, worse, or more informative
than other functions defining the same surface.

i. Note the qualifier “mathematical” in connection with the role
of \( g \), as described above. In applications the function \( g \) may
arise naturally out of the modelling, and have important inter-
pretations.

C. An Illustrative Example
1. We consider the problem

$$\max f(x, y) \text{ subject to } g(x, y) = 0$$

where $f(x, y) = x + y$ and $g(x, y) = x^2 + y^2 - 1$.

2. The gradient of $f$, at every $(x, y)$, is the vector $(1, 1)$.

3. The gradient of $g$ at $(x, y)$ is the vector $(2x, 2y)$. At each point on the constraint set, which is a circle, the gradient is a vector pointing outward.

4. Based on this geometric reasoning, we see that the only candidates for solutions are $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ and $(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$. Compactness of the constraint set, together with continuity of the objective function, implies that a solution exists, and the objective function is greater at $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ than at $(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$, so $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ is the unique solution.

5. To compute mechanically, set

$$L = x + y + \lambda(x^2 + y^2 - 1)$$

and write the necessary conditions:

$$0 = \frac{\partial L}{\partial x} = 1 + 2\lambda x;$$

$$0 = \frac{\partial L}{\partial y} = 1 + 2\lambda y;$$

$$0 = \frac{\partial L}{\partial \lambda} = x^2 + y^2 - 1.$$

Combining the first two equations gives $x = y$, and this and the third equation have the two solutions $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ and $(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$ for $(x, y)$.

III. Conceptual Interpretation

A. Not many economists know very much about the problems that Lagrange, an eighteenth century physicist, was attempting to solve.
1. A class of problems, of obvious importance in engineering applications of Newtonian physics, deals with finite collections of point-masses joined by rigid (weightless) rods.

   a. Imagine that the point masses are joints, so that the angles of the rods attached to a single mass are allowed to swivel.

2. The equilibria of such a system can very frequently be viewed as minimizing a function (e.g., potential energy) subject to the constraints imposed by the rods.

3. Lagrange’s concept was that in equilibrium there are “virtual forces” acting parallel to the various rods that, in sum at each point mass, exactly balance the external “real” forces (e.g., gravity) acting on that point mass.

   B. In economics we tend to think of the Lagrangean multiplier $\lambda$ as a “shadow price,” expressing the value, in utility or dollars or whatever $f$ measures, per unit of whatever $g$ measures, of relaxing or tightening the constraint.

IV. Multiple Constraints: Geometric Motivation

   A. The problem whose solutions we are interested in characterizing is:

   \[
   \max f(x) \quad \text{subject to} \quad g_1(x) = 0, \ldots, g_k(x) = 0,
   \]

   where $f, g_1, \ldots, g_k : \mathbb{R}^n \rightarrow \mathbb{R}$ are $C^1$ functions.

   B. As the starting point in developing the Lagrangean methods for the more general class of problems here, we recall the final picture we had of the necessary conditions in the case of a single constraint:

   The gradient of $f$ is perpendicular to the surface determined by the condition $g = 0$.

   1. Let $M = \{ x \in \mathbb{R}^n : g_1(x) = \ldots = g_k(x) = 0 \}$ be the feasible set.
a. In the case of a single constraint \((k = 1)\) we thought of \(M\) as an \((n-1)\)-dimensional surface (more precisely, a *hypersurface*). Now \(M\) will typically have lower dimension, so we might visualize \(M\) as a curve in \(\mathbb{R}^3\).

2. Let \(x^* \in M\) be a solution of the problem, and let \(v \in \mathbb{R}^n\) be a vector that is tangent to \(M\) at \(x^*\).
   a. We won’t attempt to describe precisely what we mean by a tangent vector, since it would take us far afield.
   b. Intuitively the idea is that \(v\) is a direction in which one could move, starting at \(x^*\), without leaving \(M\). The first order condition for optimization is that such motion cannot increase \(f\) or decrease \(f\) (since one could also move in the direction \(-v\)).
   c. Recall that the gradient of \(f\) at a point \(x \in \mathbb{R}^n\) is the vector
      \[
      \nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix}
      \]
      and the first order effect of moving in the direction \(v\) at \(x^*\) is
      \[
      \nabla f(x^*) \cdot v.
      \]

3. In order for \(v\) to be a vector tangent to \(M\) at \(x^*\) it is necessary that
   \[
   \nabla g_1(x^*) \cdot v = \ldots = \nabla g_k(x^*) \cdot v = 0.
   \]
The Lagrangean theorem considers the situation in which this is also a sufficient condition for \(v\) to be a vector tangent to \(M\) at \(x^*\).
   a. Example: Let \(n = 3\) and \(k = 2\) with \(g_1(x, y, z) = z - x^3\), \(g_2(x, y, z) = z - y^3\), and \(x^* = (0, 0, 0)\) then \(\nabla g_1(x^*) = \nabla g_2(x^*) = (0, 0, 1)\). However, there are many vectors of the form \((a, b, 0)\) that are not tangent to the curve \(x^3 = y^3 = z\). The problem is that the vectors \(\nabla g_1(x^*)\) and \(\nabla g_2(x^*)\) are collinear.

**Definition:** Vectors \(v_1, \ldots, v_m \in \mathbb{R}^n\) are *linearly independent* if there do not exist
\[ \alpha_1, \ldots, \alpha_m \in \mathbb{R} \text{ with some } \alpha_j \neq 0 \text{ and } \]
\[ \alpha_1 \cdot v_1 + \ldots + \alpha_m \cdot v_m = (0, \ldots, 0) \in \mathbb{R}^m. \]

b. In the example the vectors \( \nabla g_1(x^*) \) and \( \nabla g_2(x^*) \) are not linearly independent. It turns out that this is the only thing that can go wrong. The final result we are developing is

**The Lagrangean Theorem:** Suppose that \( f, g_1, \ldots, g_k : \mathbb{R}^m \to \mathbb{R} \) are \( C^1 \) functions, \( x^* \) is a solution to the problem

\[
\max f(x) \text{ subject to } g_1(x) = 0, \ldots, g_k(x) = 0,
\]

and \( \nabla g_1(x^*), \ldots, \nabla g_k(x^*) \) are linearly independent. Then there exists \( \lambda^* = (\lambda_1^*, \ldots, \lambda_k^*) \in \mathbb{R}^k \) such that

\[
(\ast) \quad \nabla f(x^*) = \lambda_1^* \nabla g_1(x^*) + \ldots + \lambda_k^* \nabla g_k(x^*).
\]

Defining

\[
\mathcal{L}(x, \lambda) = f(x) - \lambda_1 g_1(x) - \ldots - \lambda_k g_k(x),
\]

the following equations are satisfied:

(a) \( \frac{\partial \mathcal{L}}{\partial x_i}(x^*, \lambda^*) = 0 \text{ for } i = 1, \ldots, n; \)

(b) \( \frac{\partial \mathcal{L}}{\partial \lambda_j}(x^*, \lambda^*) = 0 \text{ for } j = 1, \ldots, k. \)

4. Note that (a) is simply a rephrasing of (\ast), while (b) is merely a restatement of the feasibility conditions \( g_1(x) = 0, \ldots, g_k(x) = 0 \). Therefore it is only necessary to find a \( \lambda^* \) for which (\ast) holds.

C. An Example

1. We consider the problem

\[
\max f(x, y, z) \text{ subject to } g_1(x, y, z) = 0 \text{ and } g_2(x, y, z) = 0
\]
where \( f(x, y, z) = x + y + z \), \( g_1(x, y, z) = x^2 + y^2 + z^2 - 1 \), and \( g_2(x, y, z) = z \).

2. The gradient of \( f \), at every \((x, y, z)\), is the vector \((1, 1, 1)\).

3. The gradient of \( g_1 \) at \((x, y, z)\) is the vector \((2x, 2y, 2z)\).

4. The gradient of \( g_2 \), at every \((x, y, z)\), is the vector \((0, 0, 1)\).

5. In order for \( \nabla f(x, y, z) \) to be a linear combination of \( \nabla g_1(x, y, z) \) and \( \nabla g_2(x, y, z) \) we must have \( x = y \), so the only candidates for solutions are \((\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)\) and \((-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0)\). As before, compactness of the constraint set, together with continuity of the objective function, implies that a solution exists, and the objective function is greater at \((\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)\) than at \((-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0)\), so \((\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)\) is the unique solution.

6. To compute mechanically, set

\[
\mathcal{L} = x + y + z + \lambda_1(x^2 + y^2 + z^2 - 1) + \lambda_2 z
\]

and write the necessary conditions:

\[
0 = \frac{\partial \mathcal{L}}{\partial x} = 1 + 2\lambda_1 x;
\]
\[
0 = \frac{\partial \mathcal{L}}{\partial y} = 1 + 2\lambda_1 y;
\]
\[
0 = \frac{\partial \mathcal{L}}{\partial z} = 1 + 2\lambda_1 z + \lambda_2;
\]
\[
0 = \frac{\partial \mathcal{L}}{\partial \lambda_1} = x^2 + y^2 + z^2 - 1;
\]
\[
0 = \frac{\partial \mathcal{L}}{\partial \lambda_2} = z.
\]

Combining the first two equations gives \( x = y \), and this and the last two equations have the two solutions \((\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)\) and \((-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0)\) for \((x, y, z)\).