Natalia Lazzati

Mathematics for Economics

Note 6: Nonlinear Programming - Unconstrained Optimization

Note 6 is based on de la Fuente (2000, Ch. 7), Madden (1986, Ch. 3 and 5) and Simon and Blume (1994, Ch. 17).

One objective of using economic models is to make predictions concerning the behavior of individuals and groups in situations of interest. This is possible if their behavior exhibits some sort of regularity. In economic theory it is often assumed that individuals have well-specified and consistent preferences over the set of possible results and their actions; and, that given those preferences, they choose their actions so as to obtain the best result among those available.

The latter postulates lead us to model the behavior of economic agents as the outcome of either a constrained or an unconstrained optimization problem. Notes 6-8 and 11 develop the "technology" for analyzing such problems, that is, nonlinear programming.

Nonlinear Programming (NLP)

The term NLP refers to a set of mathematical methods for characterizing the solutions to (un)constrained optimization problems. In general, the basic NLP problem can be written as

$$\max_{\mathbf{x}} \{f(\mathbf{x}) : \mathbf{x} \in U\} \tag{P}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $U \subseteq \mathbb{R}^n$. That is, we seek the value of $\mathbf{x}$ that maximizes the function f within the set U in $\mathbb{R}^n$.

In problem (P), $\mathbf{x}$ is a vector of decision variables; U is the constraint set, or the set of all the possible values that $\mathbf{x}$ can take; and f is a real-valued function, known as the objective function. A rational agent will choose an optimal plan $\mathbf{x}^*$, defined as the one that maximizes the value of the objective function f over the constraint set U.

Definition 1. We say $\mathbf{x}^*$ is a (global) maximum of f on U if it is a solution to the problem (P). That is, if $\mathbf{x}^* \in U$ and $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for all $\mathbf{x}$ in U. The point $\mathbf{x}^* \in U$ is a strict maximum if it is a maximum and $f(\mathbf{x}^*) > f(\mathbf{x})$ for all $\mathbf{x} \neq \mathbf{x}^*$ in U.

We want to find a set of necessary and sufficient conditions characterizing solutions to (P). That is, a set of conditions such that if $\mathbf{x}^*$ is a solution to the problem (P), then these conditions are satisfied by $\mathbf{x}^*$ (necessary); and a set of conditions such that if $\mathbf{x}^*$ satisfies them, then $\mathbf{x}^*$ is a solution to (P) (sufficient). Although Definition 1 describes $\mathbf{x}^*$ without using any notion of differentiability, calculus facilitates its characterization. As a consequence, most of the techniques we will study involve taking derivatives. [Do not forget however that problem (P) can be solved even if f is not differentiable. For instance, the problem $\max_{\mathbf{x}} \{f(\mathbf{x}) : \mathbf{x} \in U\}$ with $f(\mathbf{x}) = |\mathbf{x}|$ and $U = (-1,1)^n$ has a unique solution at $\mathbf{x}^* = (0, ..., 0)^T$, but the objective function is not differentiable at the null-vector.

As it is often the case, we will start with a simpler task. We will first derive necessary and sufficient conditions for local (instead of global) maxima.]

Definition 2. A point $\mathbf{x}^* \in U$ is a local maximum of f on U if there is a ball B_r around $\mathbf{x}^*$ such that $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for all $\mathbf{x} \in B_r(\mathbf{x}^*) \cap U$. The point $\mathbf{x}^* \in U$ is a strict local maximum if it is a local maximum and $f(\mathbf{x}^*) > f(\mathbf{x})$ for all $\mathbf{x} \neq \mathbf{x}^*$ in $B_r(\mathbf{x}^*) \cap U$.

Having described local maxima, we will go back to the original objective of characterizing global maxima, that is, the solutions to problem (P). To this end we will use an approach that relies on both the curvature of the objective function and the convexity of the constraint set. This approach is often called concave (or quasiconcave) programming.

In this course we consider three versions of the optimization problem that differ in terms of the way in which the feasible set is described

(a) Open constraint set, where the constraint set U is just an open set U in $\mathbb{R}^n$;

(b) Lagrange problems, where the constraint set is defined as a set of equality constraints

$$U = \{\mathbf{x} \in \mathbb{R}^n : h_1(\mathbf{x}) = a_1, ..., h_m(\mathbf{x}) = a_m\}; \text{ and}$$

(c) Kuhn-Tucker problems, where the constraint set is defined as a set of inequality constraints

$$U = \{\mathbf{x} \in \mathbb{R}^n : g_1(\mathbf{x}) \leq b_1, ..., g_k(\mathbf{x}) \leq b_k\}.$$

This note studies the first case, often called unconstrained optimization problem. Notes 7 and 8 cover the other two, respectively.

Remark. This note focuses on the maximization problem. We do this because the problem $\min_{\mathbf{x}} \{f(\mathbf{x}) : \mathbf{x} \in U\}$ can be rewritten as problem (P) by substituting $f(.)$ with $-f(.)$.

Necessary and Sufficient Conditions for Local Maxima

Consider the problem

$$\max_{\mathbf{x}} \{f(\mathbf{x}) : \mathbf{x} \in U\} \tag{P.U}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and U is an open set in $\mathbb{R}^n$. Problem (P.U) is often called unconstrained optimization problem.

Let us assume for the moment that U is an open subset of the real line. Recall that if f is C^1 and x^* is a maximum of f on U, then $f'(x^*) = 0$. Visually, if x^* is a maximum of f on U, then the tangent line to the graph of f at $[f(x^*), x^*]$ must be horizontal. A similar first order condition works for a function f of n variables.

Theorem 3. (Necessary conditions for a local maximum) Let f be a C^1 function. If $\mathbf{x}^*$ is a local maximum of f on U, then

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0 \text{ for } i = 1, ..., n.$$

Proof. Let $B = B_r(\mathbf{x}^*)$ be a ball around $\mathbf{x}^*$ in U with the property that $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for all $\mathbf{x} \in B$. Since $\mathbf{x}^*$ maximizes f on B, along each line segment through $\mathbf{x}^*$ that lies in B and that is parallel to one of the axes, f takes on its maximum value at $\mathbf{x}^*$. In other words, x_i^* maximizes the function of one variable

$$x_i \to f\left(x_1^*, ..., x_{i-1}^*, x_i, x_{i+1}^*, ..., x_n^*\right)$$

for $x_i \in (x_i^* - r, x_i^* + r)$. Apply the standard one variable maximization criterion to each of these n one-dimensional problems to conclude that

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0 \text{ for } i = 1, ..., n$$

which shows our claim. ■

Let us apply the idea to a simple example.

Example 1. To find a local maximum of $f(x_1, x_2) = -x_1^3 + x_2^3 - 9x_1x_2$, we take the partial derivatives and make them equal to zero

$$\partial f(x_1, x_2)/\partial x_1 = -3x_1^2 - 9x_2 = 0 \text{ and } \partial f(x_1, x_2)/\partial x_2 = 3x_2^2 - 9x_1 = 0.$$

Then we solve for the values of x_1 and x_2 that satisfy these two equations simultaneously. The two solutions are $(0, 0)$ and $(3, -3)$. So far, we can only say that these points are two candidates for local maxima. ▲

Stationarity at $\mathbf{x}^*$ is certainly not sufficient for $\mathbf{x}^*$ to be a local maximum. Think, for instance, in the functions $f(x) = x^2$ and $f(x) = x^3$. In both cases $f'(0) = 0$, so that 0 is a critical point but it is not a local maximum. In the first case $x^* = 0$ is a global minimum, and in the second one it is a saddle point. Before introducing sufficient conditions for local maxima, we need to extend the notion of critical point to higher dimensions.

Definition 4. An n-dimensional vector $\mathbf{x}^*$ is a critical point of a function $f : \mathbb{R}^n \to \mathbb{R}$ if $\mathbf{x}^*$ satisfies

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0 \text{ for } i = 1, ..., n.$$

The critical points in Example 1 are $(0, 0)$ and $(3, -3)$. In order to determine whether any of these points is a local maximum, we will use second order derivatives of f. The next theorem describes sufficient conditions for strict local maxima.

Theorem 5. (Sufficient conditions for a strict local maximum) Let f be a C^2 function. Suppose that $\mathbf{x}^*$ is a critical point of f on U, and that its Hessian matrix is negative definite at $\mathbf{x}^*$, i.e. $D^2 f(\mathbf{x}^*)$ is negative definite. Then $\mathbf{x}^*$ is a strict local maximum of f on U.

Remark. For strict local minimum the matrix $D^2 f(\mathbf{x}^*)$ must be positive definite.

We offer a sketch of the proof. Let $\mathbf{x}^*$ be a critical point of f on U. A C^2 function can be approximated around $\mathbf{x}^*$ by its Taylor polynomial expansion

$$f(\mathbf{x}^* + \mathbf{h}) = f(\mathbf{x}^*) + Df(\mathbf{x}^*)\mathbf{h} + \frac{1}{2}\mathbf{h}^T D^2 f(\mathbf{x}^*)\mathbf{h} + E(\mathbf{h})$$

where $E(\mathbf{h})$ is the remainder term that goes to 0 very quickly as $\mathbf{h} \to \mathbf{0}$ (see Note 2). If we ignore the latter and remember that $Df(\mathbf{x}^*) = (0, ..., 0)$ (**Why?**), then

$$f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*) \approx \frac{1}{2}\mathbf{h}^T D^2 f(\mathbf{x}^*)\,\mathbf{h}.$$

Now if $D^2 f(\mathbf{x}^*)$ is negative definite, then $\frac{1}{2}\mathbf{h}^T D^2 f(\mathbf{x}^*)\,\mathbf{h} < 0$ for all $\mathbf{h} \neq \mathbf{0}$. Which is the same as to say $f(\mathbf{x}^* + \mathbf{h}) < f(\mathbf{x}^*)$ for any small $\mathbf{h} \neq \mathbf{0}$. In other words $\mathbf{x}^*$ is a strict local maximum.

A second order necessary, but not sufficient, condition for a critical point $\mathbf{x}^*$ of a C^2 function to be a local maximum (minimum) is that $D^2 f(\mathbf{x}^*)$ be negative (positive) semi-definite (think in $f(x) = x^3$ at $x = 0$!). Therefore, if f is a C^2 function and we find that $D^2 f(\mathbf{x}^*)$ is neither positive nor negative semi-definite at a critical point $\mathbf{x}^*$, then $\mathbf{x}^*$ is a saddle-point.

Example 2. Consider the problem in Example 1. The Hessian of $f(x_1, x_2)$ is

$$D^2 f(x_1, x_2) = \begin{pmatrix} -6x_1 & -9 \\ -9 & 6x_2 \end{pmatrix}.$$

At the two critical points

$$D^2 f(0,0) = \begin{pmatrix} 0 & -9 \\ -9 & 0 \end{pmatrix} \text{ and } D^2 f(3,-3) = \begin{pmatrix} -18 & -9 \\ -9 & -18 \end{pmatrix}.$$

Since $D^2 f(0,0)$ is neither positive nor negative semi-definite, then $(0,0)$ is a saddle point. At the second critical point, the matrix $D^2 f(3,-3)$ is negative definite. (**Check it**.) Therefore, by Theorem 5, the point $(3,-3)$ is a strict local maximum of $f(x_1, x_2)$. ▲

Concavity and Optimal Solutions to Problem (P.U)

In the previous section we obtained that if f is a C^2 function, $\mathbf{x}^*$ is a critical point of f on U and the Hessian matrix at the critical point is negative definite, then $\mathbf{x}^*$ is a strict local maximum. In this section we imposse more structure on the problem to extend these ideas to optimal solutions of the initial problem.

Remark. Since concave functions are defined only on convex sets, this section assumes the set U is both open and convex.

Let us restrict attention to concave functions of a single variable, and assume that x^* is a critical point. If f is also C^1, then the tangent line to the graph of f at $[f(x^*), x^*]$ is horizontal and has height $f(x^*)$. But we know from the first derivative characterization of concave functions, that the tangent line at any point lies entirely on or above the graph. Hence $f(x^*)$ is at least as large as any other possible value of $f(x)$, and x^* must be a global maximum of the objective function in the constraint set, that is, a solution to problem (P.U). The next theorem states that this argument extends to higher dimensions.

Theorem 6. (Sufficient conditions for an optimal solution) Let f be a C^1 concave function. If there exists a point $\mathbf{x}^* \in U$ that satisfies $Df(\mathbf{x}^*) = (0, .., 0)$, then $\mathbf{x}^*$ is a global maximum of $\mathbf{x}^*$ on U, that is, a solution to problem (P.U).

Proof. Since f is C^1 and concave we know that

$$f(\mathbf{x}') + Df(\mathbf{x}')(\mathbf{x} - \mathbf{x}') \geq f(\mathbf{x}) \tag{1}$$

$\forall \mathbf{x}, \mathbf{x}' \in U$. Since (1) holds for all $\mathbf{x}' \in U$ it must be true for $\mathbf{x}' = \mathbf{x}^*$, then

$$f(\mathbf{x}^*) + Df(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq f(\mathbf{x}) \tag{2}$$

$\forall \mathbf{x} \in U$. By initial assumption, $Df(\mathbf{x}^*) = (0, .., 0)$. Since $\mathbf{x}$ was arbitrarily selected, then (2) implies that $f(\mathbf{x}^*) \geq f(\mathbf{x})$ $\forall \mathbf{x} \in U$. This completes the proof. ■

Theorem 6 characterizes the maxima of concave functions that are continuously differentiable. A related question is: How many maxima can a concave function possess? To answer this question let us analyze two examples. Think first in the function $f(x) = -x^2$. This function is **strictly concave** and it has a **unique maximum** at $x^* = 0$. Let us consider now the function

$$f(x) = \left\} \begin{array}{ll} -x^4 & \text{if } x \leq 0 \\ 0 & \text{if } x \in (0, 1) \\ -(x-1)^4 & \text{if } x \geq 1 \end{array} \right. .$$

In this second case, any x^* in $[0, 1]$ is a global maximum. Then the set of maximizers is an **infinite subset** of $\mathbb{R}$. This second function is **concave**, but **not strictly concave**.

These observations are captured by the next theorem.

Theorem 7. (**Uniqueness**) A strictly concave function $f : U \to \mathbb{R}$ with convex domain $U \subseteq \mathbb{R}^n$, cannot possess more than one global maximum. It follows that if f is strictly concave, then problem (P.U) has at most one solution.

Proof. Suppose the statement is false and let $\mathbf{x}^*$ and $\mathbf{x}^{**}$ in U be two different global maxima of the strictly concave function f. Since U is convex, then $\lambda \mathbf{x}^* + (1-\lambda)\,\mathbf{x}^{**} \in U$ for all $\lambda \in [0,1]$. By strict concavity of f

$$f\left[\lambda \mathbf{x}^* + (1-\lambda)\,\mathbf{x}^{**}\right] > \lambda f\left(\mathbf{x}^*\right) + (1-\lambda)\, f\left(\mathbf{x}^{**}\right) \text{ for all } \lambda \in (0,1).$$

Since $f\left(\mathbf{x}^*\right) = f\left(\mathbf{x}^{**}\right)$ (**Why?**), then

$$f\left[\lambda \mathbf{x}^* + (1-\lambda)\,\mathbf{x}^{**}\right] > f\left(\mathbf{x}^*\right) = f\left(\mathbf{x}^{**}\right) \text{ for all } \lambda \in (0,1).$$

and $\mathbf{x}^*$ and $\mathbf{x}^{**}$ are not global maxima, which is a contradiction. ■

In the second example, the function is not strictly concave and the number of maximizers is infinite. Nevertheless, they have a nice particular structure: the set of maximizers is a convex set. This observation is also general.

Theorem 8. (**Convex solution set**) Assume a concave function $f : U \to \mathbb{R}$ with convex domain $U \subseteq \mathbb{R}^n$, has more than one global maximum. Then the set of global maxima is a convex set, i.e. if $\mathbf{x}^*$ and $\mathbf{x}^{**}$ are two different global maxima, then $\lambda \mathbf{x}^* + (1-\lambda)\,\mathbf{x}^{**}$ is also a global maximum for all $\lambda \in [0,1]$. It follows that if f is concave and problem (P.U) has more than one solution, then the solution set is a convex set.

Proof. (Show the result.) ■

We end this note with a standard economic application to firm behavior.

Example 3. Suppose a firm uses n inputs to produce a single output. If $\mathbf{x} \in \mathbb{R}^n_+$ represents an input bundle, if $y = f(\mathbf{x})$ is the firm C^1 and strictly concave production function, and if p is the selling price of its product, then the firm revenue is $pf(\mathbf{x})$.

Let $\mathbf{w}$ denote the vector of input prices. Then the firm's profits are given by

$$\pi(\mathbf{x}) = pf(\mathbf{x}) - \mathbf{w} \cdot \mathbf{x}.$$

Assume that $f(.)$ and $\mathbf{w}$ are such that the profit-maximizing firm uses positive amounts of each input so that the profit maximizing bundle, $\mathbf{x}^*$, occurs in the interior of $\mathbb{R}^n_+$. Then by Theorem 3 the partial derivatives of $\pi(\mathbf{x})$ must be zero at the profit-maximizing $\mathbf{x}^*$

$$\frac{\partial \pi}{\partial x_i}(\mathbf{x}^*) = p\frac{\partial f}{\partial x_i}(\mathbf{x}^*) - w_i = 0 \text{ for } i = 1, 2, ...n. \tag{3}$$

In particular, the marginal revenue from using one more unit of input i must just balance the marginal cost of purchasing another unit of input i.

Notice that $\pi(\mathbf{x})$ is strictly concave on $\mathbb{R}^n_+$. (**Why?**) Then by Theorems 6 and 7 if there is an $\mathbf{x}^*$ that satisfies (3), it is the unique bundle that maximizes firm's profits. ▲