Natalia Lazzati

Mathematics for Economics

Note 7: Nonlinear Programming - The Lagrange Problem

Note 7 is based on de la Fuente (2000, Ch. 7) and Simon and Blume (1994, Ch. 18 and 19).

Introduction to the Lagrange Problem

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$, and consider the problem

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

where

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

We assume $m \leq n$, that is, the number of constraints is at most equal to the number of decision variables.[1]

This model differs from the previous one as $h_1(\mathbf{x}) = a_1, ..., h_m(\mathbf{x}) = a_m$ are m equality constraints that define the feasible set. We start by giving an intuitive interpretation of the method of Lagrange multipliers that we will use to solve this new problem.

Consider a simple version of (P.L) with only two decision variables and one constraint

$$\max_{x_1,x_2} \{f(x_1, x_2) : h(x_1, x_2) = a\}. \tag{1}$$

In Intermediate Microeconomics you learned how to solve a similar problem graphically. In Figure 1 the thicker curve is formed by the points that satisfy the constraint $h(x_1, x_2) = a$. The thinner lines are the level sets of the objective function f. The optimal solution to the problem is the point (x_1^*, x_2^*) which lies on the highest possible level set of f. Given certain convexity conditions, this point can be characterized, for some μ, by

$$Df(x_1^*, x_2^*) - \mu Dh(x_1^*, x_2^*) = (0, 0) \tag{2}$$

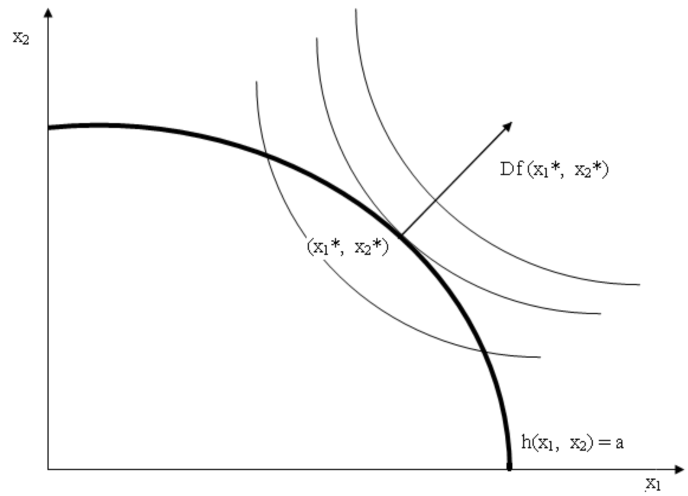
Clearly (x_1^*, x_2^*) will also satisfy the constraint

$$h(x_1^*, x_2^*) = a. \tag{3}$$

[1] All the results in this note remain valid if $f : X \rightarrow \mathbb{R}$ where X is an open set in $\mathbb{R}^n$.

The method of Lagrange multipliers transforms the constrained optimization problem (1) in an unconstrained problem that has the same solution(s).

Figure 1. Graphical Representation of the Lagrange Problem



Instead of directly forcing the agent to respect the constraint, imagine that we allow him to choose the value of the choice variables x_1 and x_2 freely, but make him pay a fine μ "per unit violation" of the restriction. The agent's payoff, net of penalty, is given by the Lagrangian function

$$L\left(x_1, x_2, \mu\right) = f\left(x_1, x_2\right) - \mu\left[h\left(x_1, x_2\right) - a\right]. \tag{4}$$

The agent then maximizes (4), taking μ as given. By Note 6, the first order necessary conditions for this problem are

$$\frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} - \mu\frac{\partial h}{\partial x_1} = 0 \text{ and } \frac{\partial L}{\partial x_2} = \frac{\partial f}{\partial x_2} - \mu\frac{\partial h}{\partial x_2} = 0. \tag{5}$$

Given an arbitrary μ there is not guarantee that the solutions to this system of equations will be optimal solutions to the original problem. However, if we pick the correct penalty μ^*, then the agent will have incentives to satisfy the constraint, and then the artificial function we have constructed will give us the correct answer. Then μ^* must be such that the constraint holds. Hence, in addition to the two equalities in (5) the solution (x_1^*, x_2^*, μ^*) must satisfy the feasibility condition, that we can conveniently express as

$$\frac{\partial L}{\partial \mu} = a - h\left(x_1, x_2\right) = 0. \tag{6}$$

We have then a system of three equations that can be solved for the optimal values of the variables of choice (x_1^*, x_2^*) and the multiplier μ^*. Moreover, these conditions coincide with the ones you learned in Intermediate Microeconomics [that is, conditions (2) and (3)]. Hence, the graphical argument also suggests that the constrained maximizer of f will be characterized by conditions (5) and (6) above, that is, by the critical points of the Lagrangian function.

This process is quite impressive. We just have reduced a constrained optimization problem in two variables into an unconstrained problem of three variables of choice. A remark can be made here. This reduction would not have worked if $\partial h/\partial x_1$ and $\partial h/\partial x_2$ were both zero at the maximizer. For this reason we will make the assumption that $\partial h/\partial x_1$ and $\partial h/\partial x_2$ are not both zero at the optimal value of the variables of choice. This restriction is often called constraint qualification. (In general you will check if there is a point in the constraint set that violates this condition. If it is the case, you will include it as part of the other candidates for local maxima.)

Necessary and Sufficient Conditions for Local Maxima

We consider now the more general problem (P.L). To study this problem we need to extend to n variables of choice and m constraint functions the qualification constraint that we used for the case of one constraint and two variables of choice.

If we have just one constraint, $h(x_1, ..., x_n) = a$, then the natural generalization of the previous constraint qualification is that the first-order partial derivatives of h are not all zero at the optimal $\mathbf{x}^*$, or

$$\left[\frac{\partial h}{\partial x_1}(\mathbf{x}^*), ..., \frac{\partial h}{\partial x_n}(\mathbf{x}^*)\right] \neq (0, ..., 0).$$

Let us define $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), ..., h_m(\mathbf{x})]^T$. If we are dealing with n variables of choice and m constraints, the new natural candidate for the qualification constraint involves the Jacobian matrix

$$D\mathbf{h}(\mathbf{x}) = \begin{pmatrix} \frac{\partial h_1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial h_1}{\partial x_n}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial h_m}{\partial x_n}(\mathbf{x}) \end{pmatrix}.$$

We say that $(h_1, ..., h_m)$ satisfies the nondegenerate constraint qualification (NDCQ) at $\mathbf{x}^*$ if the rank of the Jacobian matrix $D\mathbf{h}(\mathbf{x})$ at $\mathbf{x} = \mathbf{x}^*$ is m, that is, if $D\mathbf{h}(\mathbf{x}^*)$ has full rank. The NDCQ is a regularity condition [de la Fuente (2000), pp. 285-286, has an interesting justification for this requirement]. This condition can be violated in two ways (i) if one row has all its elements equal to zero—this violation is similar to the analyzed before; and/or (ii) if one row can be expressed as a convex combination of the other ones.

The next theorem provides necessary conditions for local maxima.

Theorem 1. (Lagrange conditions) Let $f, h_1, ..., h_m$ be C^1 functions. Suppose that $\mathbf{x}^*$ is a (local) maximum of f on U, and that $\mathbf{x}^*$ satisfies the NDCQ above. Then there exist $\mu_1^*, ..., \mu_m^*$ such that $\mathbf{x}^*$ is a critical point of the Lagrangian function

$$L(\mathbf{x}, \boldsymbol{\mu}) \equiv f(\mathbf{x}) - \mu_1^* [h_1(\mathbf{x}) - a_1] - ... - \mu_m^* [h_m(\mathbf{x}) - a_m]. \tag{7}$$

That is,

$$Df(\mathbf{x}^*) - \sum\nolimits_{i=1}^{m} \mu_i^* Dh_i(\mathbf{x}^*) = (0, ..., 0). \tag{8}$$

Remark. The Lagrange conditions imply that $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ is a critical point of $L(\mathbf{x}, \boldsymbol{\mu})$. (**Why?**).

Proof. The theorem assumes that $\mathbf{x}^*$ maximizes f on the constraint set

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

and that the Jacobian matrix

$$D\mathbf{h}(\mathbf{x}^*) = \begin{pmatrix} \frac{\partial h_1}{\partial x_1}(\mathbf{x}^*) & \cdots & \frac{\partial h_1}{\partial x_n}(\mathbf{x}^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1}(\mathbf{x}^*) & \cdots & \frac{\partial h_m}{\partial x_n}(\mathbf{x}^*) \end{pmatrix}.$$

has maximal rank m.

We first claim that the $(m+1) \times n$ Jacobian matrix

$$\begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}^*) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{x}^*) \\ \frac{\partial h_1}{\partial x_1}(\mathbf{x}^*) & \cdots & \frac{\partial h_1}{\partial x_n}(\mathbf{x}^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1}(\mathbf{x}^*) & \cdots & \frac{\partial h_m}{\partial x_n}(\mathbf{x}^*) \end{pmatrix}. \tag{9}$$

does not have maximal rank. Let $f(\mathbf{x}^*) = a_0$, and consider the system of equations

$$\begin{aligned} f(\mathbf{x}) &= a_0 \\ h_1(\mathbf{x}) &= a_1 \\ \vdots &= \vdots \\ h_m(\mathbf{x}) &= a_m. \end{aligned} \tag{10}$$

We know that $\mathbf{x}^*$ is a solution to (10). Think on the right hand side as exogenous variables. Then the matrix (9) is simply the Jacobian of the system (10) with respect to the endogenous variables $x_1, ..., x_n$.

Suppose the matrix (9) does have maximal rank $m+1$. Then by the IFT we can vary the a_i's to a_i''s a little bit and still find a solution to the revised system (10) with the a_i''s on the right hand side. In particular we can select $a_0' = a_0 + \varepsilon$ and $a_i' = a_i$ for $i = 1, ..., m$, where ε is a small positive number. By the IFT there will be a solution $(x_1^{**}, ..., x_n^{**})$ to the perturbed system

$$\begin{aligned} f(\mathbf{x}) &= a_0 + \varepsilon \\ h_1(\mathbf{x}) &= a_1 \\ \vdots &= \vdots \\ h_m(\mathbf{x}) &= a_m. \end{aligned} \tag{11}$$

Notice that $(x_1^{**}, ..., x_n^{**})$ is in U and $f(\mathbf{x}^{**}) = a_0 + \varepsilon > a_0 = f(\mathbf{x}^*)$. This is a contradiction, and therefore (9) does not have rank $m+1$. This means that its $m+1$ rows are linearly dependent; that is, there exist scalars $\alpha_0, \alpha_1, ..., \alpha_m$, not all zero, such that

$$\alpha_0 \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}^*) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}^*) \end{pmatrix} + \alpha_1 \begin{pmatrix} \frac{\partial h_1}{\partial x_1}(\mathbf{x}^*) \\ \vdots \\ \frac{\partial h_1}{\partial x_n}(\mathbf{x}^*) \end{pmatrix} + ... + \alpha_m \begin{pmatrix} \frac{\partial h_m}{\partial x_1}(\mathbf{x}^*) \\ \vdots \\ \frac{\partial h_m}{\partial x_n}(\mathbf{x}^*) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{12}$$

By the NDCQ, we conclude that $\alpha_0 \neq 0$. (**Why?**) So we can divide (12) by α_0 and let $\mu_i = -\alpha_i/\alpha_0$ for $i = 1, ..., m$ to get

$$\begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}^*) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}^*) \end{pmatrix} - \mu_1 \begin{pmatrix} \frac{\partial h_1}{\partial x_1}(\mathbf{x}^*) \\ \vdots \\ \frac{\partial h_1}{\partial x_n}(\mathbf{x}^*) \end{pmatrix} - ... - \mu_m \begin{pmatrix} \frac{\partial h_m}{\partial x_1}(\mathbf{x}^*) \\ \vdots \\ \frac{\partial h_m}{\partial x_n}(\mathbf{x}^*) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \tag{13}$$

which completes the proof. ■

The next example sheds light on the implementation of the Lagrangian method.

Example 1. Consider the problem of maximizing $f(x_1, x_2, x_3) = x_1^{1/3} x_2^{1/3} x_3^{1/3}$ subject to

$$h_1(x_1, x_2, x_3) \equiv x_1 + x_3 = 1 \text{ and } h_2(x_1, x_2, x_3) \equiv x_2^2 + x_3^2 = 1.$$

First compute the Jacobian matrix of the constraint set

$$D\mathbf{h}(\mathbf{x}) = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2x_2 & 2x_3 \end{pmatrix}.$$

Notice that its rank is 2 unless $x_2 = x_3 = 0$. Since (for any x_1) $x_2 = x_3 = 0$ violates the constraint set, then all points in the constraint set satisfy the NDCQ. The Lagrangian function is given by

$$L(x_1, x_2, x_3, \mu_1, \mu_2) = x_1^{1/3} x_2^{1/3} x_3^{1/3} - \mu_1 (x_1 + x_3 - 1) - \mu_2 \left(x_2^2 + x_3^2 - 1\right).$$

Its partial derivatives are

$$\begin{aligned}
\frac{\partial L}{\partial x_1} &= (1/3)\, x_1^{-2/3} x_2^{1/3} x_3^{1/3} - \mu_1 = 0 \\
\frac{\partial L}{\partial x_2} &= (1/3)\, x_1^{1/3} x_2^{-2/3} x_3^{1/3} - 2\mu_2 x_2 = 0 \\
\frac{\partial L}{\partial x_3} &= (1/3)\, x_1^{1/3} x_2^{1/3} x_3^{-2/3} - \mu_1 - 2\mu_2 x_3 = 0 \\
\frac{\partial L}{\partial \mu_1} &= 1 - x_1 - x_3 = 0 \\
\frac{\partial L}{\partial \mu_2} &= 1 - x_2^2 - x_3^2 = 0.
\end{aligned}$$

After the corresponding substitutions we obtain the Largangian has four critical points

$$\begin{array}{lll} x_1 = 0.4373 & x_2 = \pm 0.9008 & x_3 = 0.5657 \\ x_1 = -0.7676 & x_2 = \pm 0.6409 & x_3 = 1.7676 \end{array}.$$

So far, these are just candidates for maxima. ▲

Theorem 1 provides necessary conditions for local maxima: if $\mathbf{x}^*$ is a (local) maximum of f on U (and the NDCQ holds at $\mathbf{x}^*$) then there exist a vector $\boldsymbol{\mu}^*$ so that $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ is a critical point of the Lagrangian function. These conditions are certainly not sufficient for $\mathbf{x}^*$ to be a maximum.

In order to discriminate (among the critical points) between local maxima, local minima and saddle points we use an alternative set of sufficient conditions. These second order conditions often entail interesting information to perform comparative statics, as we will study later.

Intuitively, the second order conditions for a constrained optimization problem should involve the negative definiteness of some Hessian matrix as in the unconstrained problem, but now they should only be concerned with directions along the constraint set. The next example elaborates on this idea.

Example 2. Suppose that the objective function $f(\mathbf{x}) = \mathbf{x}^T H \mathbf{x}$ is quadratic for some symmetric matrix H, and that the constraint set is defined by the linear equations $\mathbf{h}(\mathbf{x}) \equiv A\mathbf{x} = \mathbf{0}$. Since $\mathbf{0}$ is in the constraint set and since it is a critical point of f, it is natural to ask whether $\mathbf{0}$ is the constraint maximum. Analytically, we want to know whether

$$\mathbf{x}^T H \mathbf{x} \leq 0 \text{ for all } \mathbf{x} \text{ such that } A\mathbf{x} = \mathbf{0}.$$

Since this is the same as asking whether the matrix H is negative definite on the constraint set $A\mathbf{x} = \mathbf{0}$, the answer is in Note 4. ▲

In the general problem (P.L), the first order conditions involve finding the critical points of the Lagrangian function

$$L(\mathbf{x}, \boldsymbol{\mu}) \equiv f(\mathbf{x}) - \mu_1 [h_1(\mathbf{x}) - a_1] - ... - \mu_m [h_m(\mathbf{x}) - a_m]. \tag{14}$$

Let $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ be a critical point of L. We expect that the second order conditions involve the negative definiteness of a quadratic form along a linear constraint set. A natural candidate for the quadratic form is the Hessian of the Lagrangian function with respect to $x_1, ..., x_n$ (as in the unconstrained problem). The natural linear constraint is the hyperplane that is tangent to the constraint set $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{h}(\mathbf{x}) = \mathbf{a}\}$ at the point $\mathbf{x}^*$. The next theorem shows that our intuition is, in fact, correct.

Theorem 2. (Sufficient conditions for a strict local maximum) Let $f, h_1, ..., h_m$ be C^2 functions, and assume $\mathbf{x}^*$ is a feasible point satisfying the Lagrange conditions for some $\boldsymbol{\mu}^*$.

Suppose the Hessian of the Lagrangian function (7) with respect to $\mathbf{x}$ at $(\mathbf{x}^*, \boldsymbol{\mu}^*)$, $D^2_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\mu}^*)$, is negative definite on the linear constraint set $\{\mathbf{v} : D\mathbf{h}(\mathbf{x}^*)\mathbf{v} = \mathbf{0}\}$, that is,

$$\mathbf{v} \neq \mathbf{0} \text{ and } D\mathbf{h}(\mathbf{x}^*)\mathbf{v} = \mathbf{0} \Rightarrow \mathbf{v}^T D^2_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\mu}^*)\mathbf{v} < 0.$$

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

Remark. If the goal is to minimize the objective function, replace negative definite by positive definite and $<$ by $>$ in the last restriction

Proof. See SB pp. 462-463 for the proof of Theorem 2 in the case of two variables of choice and one restriction. ■

Note 4 provides a tractable way to check the negative definiteness of a quadratic form on a linear restriction. To apply that idea to the last condition in Theorem 2, we first border the $n \times n$ Hessian matrix $D^2_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\mu}^*)$ with the $m \times n$ Jacobian matrix of the equality constraints $D\mathbf{h}(\mathbf{x}^*)$

$$H \equiv \begin{pmatrix} \mathbf{0} & D\mathbf{h}(\mathbf{x}^*) \\ D\mathbf{h}(\mathbf{x}^*)^T & D^2_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\mu}^*) \end{pmatrix} = \begin{pmatrix} 0 & \cdots & 0 & \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{\partial h_m}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_n} \\ \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_1} & \frac{\partial^2 L}{\partial x_1^2} & \cdots & \frac{\partial^2 L}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_1}{\partial x_n} & \cdots & \frac{\partial h_m}{\partial x_n} & \frac{\partial^2 L}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 L}{\partial x_n^2} \end{pmatrix}.$$

If the last $(n-m)$ leading principal minors of H alternate sign, with the sign of the determinant of matrix H the same as the sign of $(-1)^n$, then the last requirement of Theorem 2 holds.

There is a more natural way to describe the previous test. The Hessian of the Lagrangian (7) with respect to all the $(m+n)$ variables is

$$D^2 L(\mathbf{x}^*, \boldsymbol{\mu}^*) = \begin{pmatrix} 0 & \cdots & 0 & -\frac{\partial h_1}{\partial x_1} & \cdots & -\frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & -\frac{\partial h_m}{\partial x_1} & \cdots & -\frac{\partial h_m}{\partial x_n} \\ -\frac{\partial h_1}{\partial x_1} & \cdots & -\frac{\partial h_m}{\partial x_1} & \frac{\partial^2 L}{\partial x_1^2} & \cdots & \frac{\partial^2 L}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial h_1}{\partial x_n} & \cdots & -\frac{\partial h_m}{\partial x_n} & \frac{\partial^2 L}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 L}{\partial x_n^2} \end{pmatrix}.$$

Since we can obtain $D^2 L(\mathbf{x}^*, \boldsymbol{\mu}^*)$ from H by multiplying in the former the last n rows and each of the last n columns by -1, both matrices have the same principal minors. (**Why?**) Therefore we can use $D^2 L(\mathbf{x}^*, \boldsymbol{\mu}^*)$ instead of H to check the main requirement of Theorem 2; the rule does not change.

Example 3. Consider the problem in Example 1. The border Hessian H is given by

$$H \equiv \begin{pmatrix} 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 2x_2 & 2x_3 \\ 1 & 0 & (-2/9)\, x_1^{-5/3} x_2^{1/3} x_3^{1/3} & (1/9)\, x_1^{-2/3} x_2^{-2/3} x_3^{1/3} & (1/9)\, x_1^{-2/3} x_2^{1/3} x_3^{-2/3} \\ 0 & 2x_2 & (1/9)\, x_1^{-2/3} x_2^{-2/3} x_3^{1/3} & (-2/9)\, x_1^{1/3} x_2^{-5/3} x_3^{1/3} - 2\mu_2 & (1/9)\, x_1^{1/3} x_2^{-2/3} x_3^{-2/3} \\ 1 & 2x_3 & (1/9)\, x_1^{-2/3} x_2^{1/3} x_3^{-2/3} & (1/9)\, x_1^{1/3} x_2^{-2/3} x_3^{-2/3} & (-2/9)\, x_1^{1/3} x_2^{1/3} x_3^{-5/3} - 2\mu_2 \end{pmatrix}.$$

(Use the previous test to evaluate the four critical points obtained in Example 1.)

Concavity and Optimal Solutions to Problem (P.L)

In Note 6 we showed for the unconstrained optimization problem that the necessary conditions for local maxima are also sufficient if the objective function is concave. Moreover, we showed that if the objective function is concave then the local maximizers are in fact global maxima. A similar result holds for constrained optimization problems, but here we need some extra conditions on the constraint functions as well.

Theorem 3. (Sufficient conditions for an optimal solution) Let $f, -h_1, ..., -h_m$ be C^1 and concave functions. If $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ satisfies the Lagrange condition, $Df(\mathbf{x}^*) - \sum_{i=1}^m \mu_i^* Dh_i(\mathbf{x}^*) = (0, ..., 0)$, with $\mathbf{x}^*$ in U and $\boldsymbol{\mu}^* \geq (0, ..., 0)$, then $\mathbf{x}^*$ is a solution to problem (P.L).

Proof. From the first order conditions

$$Df(\mathbf{x}^*) - \sum_{i=1}^m \mu_i^* Dh_i(\mathbf{x}^*) = (0, ..., 0). \tag{15}$$

Let $\mathbf{x}$ be an arbitrary point in the constraint set U, so that $h_i(\mathbf{x}) = h_i(\mathbf{x}^*)$ for $i = 1, ..., m$. Since $-h_i$ is C^1 and concave for $i = 1, ..., m$, then

$$-Dh_i(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq 0 \text{ for } i = 1, ..., m. \text{ } (\textbf{Why?})$$

Since $\mu_1^*, ..., \mu_m^*$ are all positive, then

$$-\sum_{i=1}^m \mu_i^* Dh_i(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq 0.$$

By (15) it follows that

$$Df(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \leq 0. \text{ } (\textbf{Why?})$$

Since f is C^1 and concave

$$f(\mathbf{x}^*) \geq f(\mathbf{x}). \; (\textbf{Why?})$$

The result follows as $\mathbf{x}$ is an arbitrary point in the constraint set U. ■