The Role of the Implicit Function in the Lagrange Multiplier Rule

E. Bendito, A. Carmona, A.M. Encinas
Departament de Matemàtica Aplicada III
Universitat Politècnica de Catalunya

Abstract

In this note we present the necessary condition of constrained extremum point and the criteria for its classification using the implicit function, the chain rule and elementary techniques of linear algebra.

Key words. Multiplier rule, Implicit Function Theorem.

AMS subject classification. 2601, 26B10

The constrained extremum problem is a topic that appears in most calculus courses. However it seems to be difficult to decide the kind of critical point. We may find two types of textbooks in the literature:
• Those that introduce the Lagrange Function to give the necessary condition but do not explain the sufficient condition because its proof exceeds their scope. (See [1, 2, 5].)
• Those that give the necessary condition by means of the Lagrange Function and, to obtain the sufficient condition, they introduce the concept of differential manifold and its tangent space. (See [3, 4].)

Our purpose is to show that the constrained extremum problem is a natural extension of the unconstrained case, and therefore it can be handled in a similar way. Firstly, recall that an unconstrained extremum problem can be written as follows:

Given $\Omega_n \subset \mathbb{R}^n$ an open set and $f : \Omega_n \rightarrow \mathbb{R}$ a $C^2$-function, obtain the local extrema of $f$.

Both, the necessary and sufficient conditions for this problem are well known. (See [1].)

On the other hand, a constrained extremum problem can be written as follows:
Given $\Omega_n \subset \mathbb{R}^n$ an open set, $f : \Omega_n \to \mathbb{R}$ and $g : \Omega_n \to \mathbb{R}^m$ $C^2$-functions such that $\text{rank}(Dg(y)) = m < n$, $\forall y \in \Omega_n$ and $M = \{y \in \Omega_n : g(y) = 0\}$, obtain the local extrema of $f|_M$.

How can we handle this problem? Why must we introduce the Lagrange function?

A possible answer to the last question could be because the constrained extremum problem is transformed into an unconstrained extremum problem. Let us examine the Lagrange function which is defined as:

$$L : \Omega_n \times \mathbb{R}^m \to \mathbb{R}$$

$$L(y, \lambda) = f(y) - \lambda_1 g_1(y) - \cdots - \lambda_m g_m(y).$$

We can see that this function depends on $m$ new variables, so the dimension of the underlying space has been increased. It would be desirable that this increase would allow us to apply the techniques of unconstrained extremum classification, but this is not possible as the following simple example shows:

The function $f(y_1, y_2) = y_1^2 + y_2^2$ attains a local minimum on $M = \{(y_1, y_2) \in \mathbb{R}^2 : y_2 = 1\}$ at the point $(0, 1)$ with multiplier $\lambda = -2$. Nevertheless, the point $(0, 1, -2)$ is a saddle point of function $L$.

Hence, we cannot act the same way as in the unconstrained case. Maybe, it would be good to reflect on the initial problem, that is, to obtain the local extrema of $f$ on $M = \{y \in \Omega_n : g(y) = 0\}$. The difference between this problem and the unconstrained one is that the set $M$ is a closed differential manifold instead of an open set of $\mathbb{R}^n$. Is it then necessary to introduce the differential manifolds theory to classify the critical points of a constrained extremum problem? The answer is negative, because it suffices to consider the implicit function that describes $M$ from an open set of $\mathbb{R}^n$. Let us show how to do it.

Let $y_0 \in \Omega_n$. As $\text{rank}(Dg(y_0)) = m < n$, we can apply the Implicit Function Theorem to $g(y) = 0$. Then, if $x = (y_1, \ldots, y_k)$, $k = n - m$, there exist an open set $\Omega_k \subset \mathbb{R}^k$ and a $C^2$-function $\varphi : \Omega_k \to \Omega_n$ such that

$$\varphi(x) = (x, y_{k+1}(x), \ldots, y_n(x)), \quad \text{with} \quad \varphi(x_0) = y_0 \quad (1)$$

and $g \circ \varphi \equiv 0$. Hence,

$$f \circ \varphi = f|_M.$$

Then, the problem of finding the local extrema of $f|_M$ is equivalent to the problem of finding the local extrema of $f \circ \varphi : \Omega_k \to \mathbb{R}$ with the additional information given by the identity $g \circ \varphi \equiv 0$.

Consequently, we have achieved our purpose, since the constrained extremum problem can be treated as an unconstrained one. In addition, the dimension of the underlying space is smaller, which is natural, since $g(y) = 0$ defines some relations between the variables.
We must note that the above mentioned references arrive to this point. However, they leave this way in favour of the Lagrange Function. We will keep the way and will proceed applying the unconstrained extremum theory.

Let us make an abstraction of the role played by the implicit function to enlighten the problem.

Study the local extrema of the function \( f \circ h \), knowing that \( g \circ h \equiv 0 \), where \( f : \Omega_n \rightarrow \mathbb{R} \), \( g : \Omega_n \rightarrow \mathbb{R}^m \), and \( h : \Omega_k \rightarrow \mathbb{R}^n \) are \( C^2 \)-functions with \( \Omega_n \subset \mathbb{R}^m \), \( \Omega_k \subset \mathbb{R}^k \) open sets and \( h(\Omega_k) \subset \Omega_n \).

Let \( x_0 \in \Omega_k \) be a point where a extremum of \( f \circ h \) occurs such that \( \text{rank}(Dg(y_0)) = n - \text{rank}(Dh(x_0)) \), where \( y_0 = h(x_0) \).

It will be useful to consider the transposed Jacobian matrices of \( h \) and \( g \) denoted by 
\[
A = (\nabla h_1(x_0), \ldots, \nabla h_n(x_0)) \quad B = (\nabla g_1(y_0), \ldots, \nabla g_m(y_0)),
\]
respectively, and 
\[V = \text{span}\{\nabla g_1(y_0), \ldots, \nabla g_m(y_0)\}.\]

By applying the chain rule and using that \( g \circ h \equiv 0 \), we obtain 
\[Dg(y_0) \circ Dh(x_0) = 0.\]

If we transpose this expression we get 
\[A B = 0,
\]
and hence \( V \subset \text{Ker}A \). As \( \text{dim}V = \text{rank}(B) = n - \text{rank}(A) = \text{dim Ker}A \), we arrive to 
\[V = \text{Ker}A. \quad (2)\]

Let us examine the necessary condition. If \( f \circ h \) attains an extremum at the point \( x_0 \), then \( D(f \circ h)(x_0) = 0 \). Applying again the chain rule we obtain 
\[Df(y_0) \circ Dh(x_0) = 0.\]

If we transpose this expression we get 
\[A \nabla f(y_0) = 0
\]
which means that \( \nabla f(y_0) \in \text{Ker}A \). Hence, by (2) there exists \( \lambda \in \mathbb{R}^m \) such that 
\[\nabla f(y_0) = \sum_{l=1}^{m} \lambda_l \nabla g_l(y_0). \quad (3)\]

Note that \( \lambda \) is not unique, except when \( \text{rank}(Dg(y_0)) = m \).
To derive the sufficient condition, we must know $H(f \circ h)$, the Hessian matrix of $f \circ h$.

\[ D_{ij}(f \circ h)(x_0) = \sum_{p,q=1}^{n} D_{pq}f(y_0)D_{ij}h_q(x_0)D_ih_p(x_0) + \sum_{p=1}^{n} D_{ip}f(y_0)D_{ij}h_p(x_0). \] (4)

Analogously, for all $l = 1, \ldots, m$,

\[ D_{ij}(g_l \circ h)(x_0) = \sum_{p,q=1}^{n} D_{pq}g_l(y_0)D_{ij}h_q(x_0)D_ih_p(x_0) + \sum_{p=1}^{n} D_{ip}g_l(y_0)D_{ij}h_p(x_0) = 0. \] (5)

From (3) and (5) we get,

\[ \sum_{p=1}^{n} D_{ip}f(y_0)D_{ij}h_p(x_0) = -\sum_{l=1}^{m} \sum_{p,q=1}^{n} \lambda_l D_{pq}g_l(y_0)D_{ij}h_q(x_0)D_ih_p(x_0). \] (6)

Replacing the last term of (4) by the right hand side of equation (6) we obtain,

\[ D_{ij}(f \circ h)(x_0) = \sum_{p,q=1}^{n} D_{pq}f(y_0)D_{ij}h_q(x_0)D_ih_p(x_0). \]

Finally,

\[ H(f \circ h)(x_0) = A H(f - \sum_{l=1}^{m} \lambda_l g_l(y_0)) A^T. \] (7)

Let $q$ and $Q$ be the quadratic forms on $\mathbb{R}^k$ and $\mathbb{R}^m$ defined by

\[ q(w) = < H(f \circ h)(x_0)w, w > \quad \text{and} \quad Q(v) = < H(f - \sum_{l=1}^{m} \lambda_l g_l(y_0))v, v >, \]

respectively. Then from (7),

\[ q(w) = < H(f \circ h)(x_0)w, w > = < H(f - \sum_{l=1}^{m} \lambda_l g_l(y_0))A^T w, A^T w >= Q(A^T w) \]

and hence, $q = Q_{|\text{Im}A^T}$. On the other hand, from (2) and taking into account that $(\text{Ker}A)^\perp = \text{Im}A^T$, we get $\text{Im}A^T = V^\perp$ and therefore, $q = Q_{|V^\perp}$. Hence, $H(f \circ h)(x_0)$ is definite [strictly definite] if $H(f - \sum_{l=1}^{m} \lambda_l g_l(y_0))$ is definite [strictly definite] on $V^\perp$.

So, applying the standard criterium for classification of unconstrained extrema we have that: If the function $f \circ h$ attains a local extremum at $x_0$, then $H(f - \sum_{l=1}^{m} \lambda_l g_l(y_0))$ is definite on $V^\perp$. Moreover, if $H(f - \sum_{l=1}^{m} \lambda_l g_l(y_0))$ is strictly definite on $V^\perp$, then the function $f \circ h$ attains a local extremum at $x_0$. 4
Note that the function $h$ has disappeared in both necessary and sufficient conditions.

An immediate consequence of the above results is the Lagrange Multiplier Rule to the solution of the constrained extremum problem. For this, it suffices to take as $h$ the function, $\phi$ defined in (1).

**Lagrange Multiplier Rule.** Let $y_0 \in \Omega_n$ such that $\text{rank}(Dg(y_0)) = m$. If the function $f$ attains a local extremum on $M$ at the point $y_0$, then there exists a unique $\lambda^* \in \mathbb{R}^m$ such that 

$$\nabla f(y_0) = \sum_{i=1}^{m} \lambda_i \nabla g_i(y_0).$$

Moreover we have obtained a criterion for the classification of such critical points: If the function $f$ attains a local extremum on $M$ at the point $y_0$, then $H_y(\mathcal{L})(y_0, \lambda^*)$ is definite on $\text{span}\{\nabla g_1(y_0), \ldots, \nabla g_m(y_0)\}^\perp$. Moreover, if $H_y(\mathcal{L})(y_0, \lambda^*)$ is strictly definite on $\text{span}\{\nabla g_1(y_0), \ldots, \nabla g_m(y_0)\}^\perp$ then $f$ attains a local extremum on $M$ at the point $y_0$, where $\mathcal{L}$ is the Lagrange Function.

We must note that the Lagrange Function appears here as a consequence of the reasoning and not as an a priori construction.

To sum up, the role played by the implicit function in the constrained extremum problem is to turn it into an unconstrained one. Moreover, the additional functional information $g \circ \varphi \equiv 0$, allows us to obtain the successive derivatives of $f \circ h$ independently of the knowledge of the function $\varphi$. Therefore, the Implicit Function Theorem has been used here only once. In contrast, those that classify constrained extrema need again the Implicit Function Theorem to introduce the concept of differential manifold.

Finally, in the language of differential manifolds, we get from (7) that

$$H(f|_M)(y_0) = H_y(\mathcal{L})(y_0, \lambda^*)|_{Ty_0(M) \times Ty_0(M)}.$$ 

This result was already obtained in [6].

**References**


