Ciencia y Tecnología, 26(1 y 2): 87-100, 2010
ISSN: 0378-0524

ON THE CLASSIFICATION OF CONSTRAINED EXTREMA

*Héctor Figueroa, Orietta Protti**

Escuela de Matemáticas, Universidad de Costa Rica,
San Pedro, Costa Rica

Abstract

Using simple arguments, accessible to students of advanced calculus with an interest in Mathematics, we show the equivalence of several criteria, scattered in the literature, to classify the critical points of functions of two or three variables when restricted to side conditions.

Resumen

Usando argumentos elementales, accesibles aún para estudiantes de cálculo avanzado con un cierto interés en Matemáticas, mostramos la equivalencia de varios criterios, esparcidos en la literatura, para clasificar los puntos críticos de funciones de dos o tres variables cuando están sujetas a condiciones, o ligaduras.

Key words: Lagrange multipliers, extrema with constraints, quadratic forms.

Palabras claves: Multiplicadores de Lagrange, extremos con condiciones, formas cuadráticas.

I. Introduction

One can hardly overestimate the usefulness of optimization in all subjects in which mathematics is applied. With this in mind, we feel that a proper understanding of methods to classify the extrema of functions subject to side conditions should be included in any typical calculus, or real analysis book. However, in most books the subject is completely left aside, or briefly discussed with a barely convincing "hand waving" argument. A likely reason for this is that some general criteria are a bit hard to apply, others are somewhat cumbersome to state, and, worse, they seem unrelated at first sight. Thus, in this paper, we concentrate on functions of two or three variables where matters can be clarified using simple arguments,

* Autor para correspondencia e-mail: orieta.protti@ucr.ac.cr

with mathematics from the toolkit of an advanced calculus student. This, therefore, is an expository paper wherein we focus on the ideas, rather than the technicalities necessary to gain generality; we state several reformulations of Lagrange's criterion to determine the nature of extrema in the presence of side conditions, and prove that they are equivalent.

Even though the method of Lagrange multipliers to obtain the extrema of functions of several variables, subject to some constraints, can be proved from geometrical considerations, algebraic arguments, or analytic reasoning, for the sake of completeness, and to set our notation, here we use linear algebra ideas, as found in classic texts such as [6, Section 3.61], to establish Lagrange's technique. In what follows, we assume that all functions belong to C^3 in their domain, so all the partial derivatives up to order three exist and are continuous in their domain.

Suppose that we seek the extreme points of the function $w = f(x, y, z)$ on the curve given by the equations

$$\begin{cases} \phi^1(x, y, z) = 0, \\ \phi^2(x, y, z) = 0. \end{cases} \tag{1.1}$$

For the Lagrange method to work, we need to assume that the functions ϕ^1 and ϕ^2 are independent in some sense at a critical point $\vec{x}_0 := (x_0, y_0, z_0)$ of f subject to these constraints. One usually requires the vectors $\nabla\phi^1(\vec{x}_0)$ and $\nabla\phi^2(\vec{x}_0)$ to be linearly independent. To be concrete, we assume that, from the equations (1.1), we can solve, for instance, for x and y in terms of z; according to the implicit function theorem, it is enough to assume that the Jacobian $\frac{\partial(\phi^1,\phi^2)}{\partial(x,y)}$ does not vanish in a neighborhood of $\vec{x}_0$. A necessary condition for $\vec{x}_0$ to be an extremum of the problem is thus

$$0 = \frac{dw}{dz}(z_0) = f_x(\vec{x}_0)\frac{dx}{dz}(z_0) + f_y(\vec{x}_0)\frac{dy}{dz}(z_0) + f_z(\vec{x}_0).$$

Differentiating (1.1) with respect to z yields

$$0 = \phi^1_x(\vec{x}_0)\frac{dx}{dz}(z_0) + \phi^1_y(\vec{x}_0)\frac{dy}{dz}(z_0) + \phi^1_z(\vec{x}_0),$$
$$0 = \phi^2_x(\vec{x}_0)\frac{dx}{dz}(z_0) + \phi^2_y(\vec{x}_0)\frac{dy}{dz}(z_0) + \phi^2_z(\vec{x}_0).$$

These three equations show that the linear system $A\vec{x} = 0$, where A is the matrix

$$A = \begin{pmatrix} f_x(\vec{x}_0) & f_y(\vec{x}_0) & f_z(\vec{x}_0) \\ \phi^1_x(\vec{x}_0) & \phi^1_y(\vec{x}_0) & \phi^1_z(\vec{x}_0) \\ \phi^2_x(\vec{x}_0) & \phi^2_y(\vec{x}_0) & \phi^2_z(\vec{x}_0) \end{pmatrix},$$

has a nontrivial solution, so that $\det A = 0$. Since the linear system

$$\begin{aligned} f_x(\vec{x}_0)\, a + \phi^1_x(\vec{x}_0)\, b + \phi^2_x(\vec{x}_0)\, c = 0 \\ f_y(\vec{x}_0)\, a + \phi^1_y(\vec{x}_0)\, b + \phi^2_y(\vec{x}_0)\, c = 0 \\ f_z(\vec{x}_0)\, a + \phi^1_z(\vec{x}_0)\, b + \phi^2_z(\vec{x}_0)\, c = 0, \end{aligned} \tag{1.2}$$

has the form $B\vec{x} = 0$, with $B = A^t$, and so $\det B = \det A = 0$. In particular, the system (1.2) has also non-trivial solutions (a, b, c). Furthermore, since $\frac{\partial(\phi^1,\phi^2)}{\partial(x,y)}(\vec{x}_0) \neq 0$ it follows that $a \neq 0$. Thus, there exist real numbers λ and γ such that

$$\begin{aligned} f_x(\vec{x}_0) + \lambda\,\phi^1_x(\vec{x}_0) + \gamma\,\phi^2_x(\vec{x}_0) &= 0 \\ f_y(\vec{x}_0) + \lambda\,\phi^1_y(\vec{x}_0) + \gamma\,\phi^2_y(\vec{x}_0) &= 0 \\ f_z(\vec{x}_0) + \lambda\,\phi^1_z(\vec{x}_0) + \gamma\,\phi^2_z(\vec{x}_0) &= 0. \end{aligned} \tag{1.3}$$

The conclusion is that the extrema of our problem are either critical points of the *associated Lagrange function*

$$F(x, y, z; \lambda, \gamma) := f(x, y, z) + \lambda\,\phi^1(x, y, z) + \gamma\,\phi^2(x, y, z);$$

that is, points that satisfy (1.1) and (1.3), or points where ϕ^1 and ϕ^2 are not independent, which are, then, singular points of the curve (1.1).

Similar considerations hold for a function of n variables when restricted to m conditions, however to obtain a square matrix A we require $m < n$. This is certainly the case when there is only one restriction, and the function is either of two or three variables, which, together with the case treated above, are the usual cases considered in calculus courses. A detailed analytical proof of Lagrange's method to find the critical points of functions subject to constraints can be found in any standard textbook such as [1] or [2].

II Classification of extrema of functions of two variables on a curve

Lagrange provided the following criterion to classify extrema of functions subject to side conditions [9]:

Theorem 2.1. *Let $\vec{x}_0 \in \mathbb{R}^n$ be a critical point of a smooth function $w = f(\vec{x})$ subject to the constraints*

$$\begin{cases} \phi^1(\vec{x}) = 0, \\ \quad\vdots \\ \phi^m(\vec{x}) = 0, \end{cases} \tag{2.1}$$

where $m < n$ and $\phi^1, \ldots, \phi^m$ are also smooth. Assume that $\nabla\phi^1(\vec{x}_0), \ldots, \nabla\phi^m(\vec{x}_0)$ are linearly independent vectors in $\mathbb{R}^n$, and that F is the associated Lagrange function.

i-) *If the Hessian $HF(\vec{x}_0)(\vec{h}) > 0$, at those points $\vec{h} \neq 0$ that satisfy the linear system of equations*

$$\begin{cases} \nabla\phi^1(\vec{x}_0) \cdot \vec{h} = 0, \\ \quad\vdots \\ \nabla\phi^m(\vec{x}_0) \cdot \vec{h} = 0, \end{cases} \tag{2.2}$$

then f attains a minimum at $\vec{x}_0$.

ii-) *If $HF(\vec{x}_0)(\vec{h}) < 0$, at those points $\vec{h} \neq 0$ satisfying (2.2), then f attains a maximum at $\vec{x}_0$.*

iii-) *If $HF(\vec{x}_0)$ takes strictly positive values at some points $\vec{h}$ satisfying (2.2) and strictly negative values at other points of the same set, then f does not have an extremum at $\vec{x}_0$.*

Thus, the crux of the matter is to determine the sign of the quadratic form $HF(\vec{x}_0)$ on a certain subset of $\mathbb{R}^n$. As in the case with no constraints, the several reformulations of Lagrange's criterion are basically divided into two types: those that accomplish this task using determinants, much in the style of using Sylvester's criterion to determine when a quadratic form is definite, and those that use other, somewhat less methodical, algebraic manipulations, such as those that compress to the constrained manifold. It is likely that these criteria are cast into oblivion, or simply ignored, because they are not easy to apply in concrete examples, especially in four or higher dimensions. Furthermore, some reformulations are cumbersome even to state, as one must keep track of several signs, depending on the dimension, and the number of constraints.

To motivate the whole enterprise, we start by considering the simplest case, that is, when $\vec{x}_0 = (x_0, y_0)$ is a critical point of the Lagrange function associated with the function $w = f(x, y)$, restricted to a curve $\phi(x, y) = 0$. The first idea would be to use this equation to solve one variable in terms of the other, to consider f as a function of just one variable, and to use the test for such functions. In practice, however, this is usually difficult, as $\phi(x, y) = 0$ might be a complicated equation. Lagrange proposed instead to use Taylor's formula:

$$F(x_0 + h, y_0 + k; \lambda) \approx F(x_0, y_0; \lambda) + \nabla F(x_0, y_0; \lambda) \cdot (h, k) + \tfrac{1}{2} HF(x_0, y_0; \lambda)\,(h, k), \qquad (2.3)$$

where the symbol $\approx$ here means that one is assuming the error to be small enough to be neglected, and HF denotes the Hessian, or second differential, of the associated Lagrange function F. As $\vec{x}_0$ is an extremum, and both (x_0, y_0) and $(x_0 + h, y_0 + k)$ are points on the curve, (2.3) becomes

$$\begin{aligned} f(x_0 + h, y_0 + k) - f(x_0, y_0) &\approx \tfrac{1}{2} HF(x_0, y_0; \lambda)\,(h, k) \\ &:= \frac{1}{2}\Big(F_{xx}(\vec{x}_0)\, h^2 + F_{yy}(\vec{x}_0)\, k^2 + 2F_{xy}(\vec{x}_0)\, hk\Big). \end{aligned} \qquad (2.4)$$

Lagrange noticed that as one is interested only in those points $(x_0 + h, y_0 + k)$ that lie on the curve $\phi(x, y) = 0$, we should consider the condition

$$\nabla\phi(x_0, y_0) \cdot (h, k) = 0. \qquad (2.5)$$

At first sight, it might seen unnatural to consider (2.5), even though it is the equation of the tangent line of the curve at $\vec{x}_0$, and intuitively we should consider only such tangent directions. The point is that Lagrange [9], to develop his method, applied Taylor's theorem to the side condition $\phi(x_0 + h, y_0 + k) = 0$ to obtain

$$0 = \phi_x(x_0, y_0)\, h + \phi_y(x_0, y_0)\, k + \psi_2(h, k), \qquad (2.6)$$

in which ψ_2 denotes the terms of second and higher order. Furthermore, using the fact that the curve $\phi(x, y) = 0$ is not singular at $\vec{x}_0$, he assumed that $\phi_x(\vec{x}_0) \neq 0$, and used equation (2.6) to write h as a function of k, to study the difference $f(x_0+h, y_0+k) - f(x_0, y_0)$

as a function of the variable k. In those days, to solve an equation, or a system of equations, like (2.6), it was customary to start by setting the linear terms equal to zero; next, the approximate solution so obtained was plugged into the remaining terms to consider again linear terms of the new functions, and so on. Equation (2.5) is precisely the linear term of (2.6). This result explains the appearance of the tangent directions (2.2) of the manifold determined by (2.1) when deciding the nature of a critical point of a problem with constraints.

A detailed modern proof of Lagrange's criterion can be found, for instance, in [3]. But for the reader's convenience, we shall provide elementary proofs for the cases of functions of two or three variables.

To reformulate Theorem 2.1, in the case under discussion, notice that if $\phi_x \neq 0$ (to lighten the notation, from now on, all partial derivatives are understood to be evaluated at the critical point $\vec{x}_0$), then (2.5) entails

$$h = -\frac{\phi_y}{\phi_x}\, k,$$

and, since $HF(\vec{x}_0)$ is homogeneous of degree two, the right hand side of (2.4) is

$$\begin{aligned} HF(\vec{x}_0)(h,k) &= k^2\,(\phi_x)^{-2}\, HF(\vec{x}_0)\,(-\phi_y,\phi_x) \\ &= k^2\,(\phi_x)^{-2}\left[F_{xx}\,(\phi_y)^2 + F_{yy}\,(\phi_x)^2 - 2F_{xy}\phi_x\phi_y\right] \\ &= -k^2\,(\phi_x)^{-2}\; M, \end{aligned} \tag{2.7}$$

in which M is the determinant

$$M := \begin{vmatrix} F_{xx} & F_{xy} & \phi_x \\ F_{xy} & F_{yy} & \phi_y \\ \phi_x & \phi_y & 0 \end{vmatrix}.$$

When $\phi_x = 0$, (2.5) implies $k = 0$, since $\nabla\phi(\vec{x}_0) \neq 0$, and the right hand side of (2.4) becomes

$$HF(x_0,y_0)(h,k) = F_{xx}\,h^2 = -h^2\,(\phi_y)^{-2}\; M.$$

The sign of $HF(x_0,y_0)(h,k)$, on those points at which $\nabla\phi(\vec{x}_0)\cdot(h,k) = 0$, is thus determined by the **Hessian determinant** M, also called the **bordered determinant**. Note that the tangent space, in the case under consideration, is one-dimensional, and M is precisely the value of the Hessian $HF(x_0,y_0)$ at the point $\big(-\phi_y(\vec{x}_0),\phi_x(\vec{x}_0)\big)$, which satisfies (2.5). Unsurprisingly, then, Lagrange's criterion, for the case at hand, is equivalent to a criterion in terms of the bordered determinant M. In the following theorem we state this reformulation and prove it using elementary arguments; indeed, based on the key observation (2.7), we prove at the same time Lagrange's criterion.

Theorem 2.2. *If $\vec{x}_0$ is a critical point of the Lagrange function associated to $w = f(x,y)$ and the side condition $\phi(x,y) = 0$, and $\nabla\phi(\vec{x}_0) \neq 0$, then*

i-) *f attains a minimum at $\vec{x}_0$ when $M < 0$,*

ii-) *f attains a maximum at $\vec{x}_0$ when $M > 0$,*

iii-) *when $M = 0$, further analysis is needed to classify the critical point $\vec{x}_0$.*

Proof. Without loss of generality we assume that $\phi_x \neq 0$. By the implicit function theorem, there exist a function $x = g(y)$ satisfying $\phi\big(g(y), y\big) = 0$, in a neighborhood of y_0, then

$$0 = \phi_x \frac{dg}{dy} + \phi_y\,, \quad \text{so} \quad \frac{dg}{dy} = -\frac{\phi_y}{\phi_x}\,,$$

and

$$\frac{d^2 g}{dy^2} = \frac{d}{dy}\left(\frac{dg}{dy}\right) = \left[2\phi_{yx}\phi_y - \phi_{yy}\phi_x - \phi_{xx}\left(\phi_y\right)^2\left(\phi_x\right)^{-1}\right]\left(\phi_x\right)^{-2}.$$

To classify the critical point y_0 of the function $w(y) = f\big(g(y), y\big)$, we need to compute its second derivative

$$\begin{aligned}\frac{d^2 w}{dy^2} = \frac{d}{dy}\left(\frac{dw}{dy}\right) &= f_{xx}\left(\frac{dg}{dy}\right)^2 + 2f_{yx}\frac{dg}{dy} + f_x\frac{d^2 g}{dy^2} + f_{yy} \\ &= \left[f_{xx}\left(\phi_y\right)^2 - 2f_{yx}\phi_y\phi_x + f_{yy}\left(\phi_x\right)^2 + 2\frac{f_x}{\phi_x}\phi_{yx}\phi_y\phi_x \right. \\ &\qquad \left. - \frac{f_x}{\phi_x}\phi_{yy}\left(\phi_x\right)^2 - \frac{f_x}{\phi_x}\phi_{xx}\left(\phi_y\right)^2\right]\left(\phi_x\right)^{-2}.\end{aligned}$$

Since $\vec{x}_0$ is a critical point of the associated Lagrange's function, $f_x(\vec{x}_0) + \lambda\phi_x(\vec{x}_0) = 0$, in particular $\lambda = -f_x(\vec{x}_0)/\phi_x(\vec{x}_0)$, therefore

$$\begin{aligned}\frac{d^2 w}{dy^2} &= \Big[\left(f_{xx} + \lambda\phi_{xx}\right)\left(\phi_y\right)^2 + \left(f_{yy} + \lambda\phi_{yy}\right)\left(\phi_x\right)^2 \\ &\qquad - 2\left(f_{yx} + \lambda\phi_{yx}\right)\phi_y\phi_x\Big]\left(\phi_x\right)^{-2} \\ &= \left[F_{xx}\left(\phi_y\right)^2 + F_{yy}\left(\phi_x\right)^2 - 2F_{yx}\phi_y\phi_x\right]\left(\phi_x\right)^{-2} \\ &= \left(\phi_x\right)^{-2} HF(\vec{x}_0)\left(-\phi_y, \phi_x\right) = -\left(\phi_x\right)^{-2} M,\end{aligned}$$

upon using (2.7). Thus, the theorem follows from the second derivative test for the function $w(y) = f\big(g(y), y\big)$. □

To illustrate the lack of a criterion when $M = 0$, we analyze the following examples. Consider the function $f(x, y) = (x + y)^4$ on the circle $x^2 + y^2 = 2$. The critical points are the solutions to the system of equations

$$\begin{aligned} 4(x+y)^3 + 2\lambda x &= 0, \\ 4(x+y)^3 + 2\lambda y &= 0, \\ x^2 + y^2 - 2 &= 0, \end{aligned} \tag{2.8}$$

which, turn out to be $(1, -1; \lambda = 0)$, $(-1, 1; \lambda = 0)$, $(1, 1)$ and $(-1, -1)$. In this case,

$$M = -48(x-y)^2(x+y)^2 - 8\lambda(x^2 + y^2),$$

so M vanishes at the first two points. Moreover,

$$f(1,-1) = f(-1,1) = 0 \leq f(x,y);$$

therefore f attains a minimum at each of these points.

By contrast, if $f(x,y) = (x+y)^3$, the critical points of f, when restricted to the same circle, are obtained by replacing $4(x+y)^3$ by $3(x+y)^2$ in (2.8), therefore the solutions are exactly the same as before. Moreover,

$$M = -24(x-y)^2(x+y) - 8\lambda(x^2+y^2),$$

so, again, $M(1,-1;\lambda=0) = 0 = M(-1,1;\lambda=0)$. Once more, $f(1,-1) = f(-1,1) = 0$. Nevertheless, f is positive on the half plane to the right of the line $y=-x$, and negative on the left half plane. Thus, f attains neither a maximum nor a minimum at these two points.

III. Classification of extrema of functions of three variables on a surface

Assume now that $\vec{x}_0 := (x_0, y_0, z_0)$ is a critical point of the function $w = f(x,y,z)$ on the surface $\phi(x,y,z) = 0$. On account of Lagrange's criterion, to ascertain the nature of $\vec{x}_0$ we need to determine the sign of $HF(x_0, y_0, z_0; \lambda)(h,k,l)$, at those points (h,k,l) such that $\nabla\phi(\vec{x}_0)\cdot(h,k,l) = 0$. As we are assuming that $\nabla\phi(\vec{x}_0) \neq 0$, we may suppose, for instance, that $\phi_x(\vec{x}_0) \neq 0$. Then

$$h = -\frac{\phi_y}{\phi_x}\,k - \frac{\phi_z}{\phi_x}\,l,$$

and it is easy to verify that

$$HF(x_0,y_0,z_0)(h,k,l) = -\left(\phi_x\right)^{-2}\left[M_2\,k^2 + N_2\,l^2 + 2L_2\,kl\right], \tag{3.1}$$

where

$$\begin{aligned} M_2 &:= 2F_{xy}\phi_x\phi_y - F_{xx}\left(\phi_y\right)^2 - F_{yy}\left(\phi_x\right)^2, \\ N_2 &:= 2F_{xz}\phi_x\phi_z - F_{xx}\left(\phi_z\right)^2 - F_{zz}\left(\phi_x\right)^2, \\ L_2 &:= F_{xy}\phi_x\phi_z + F_{xz}\phi_x\phi_y - F_{yz}\left(\phi_x\right)^2 - F_{xx}\phi_y\phi_z\,. \end{aligned} \tag{3.2}$$

In particular, notice that

$$M_2 = \begin{vmatrix} F_{xx} & F_{xy} & \phi_x \\ F_{xy} & F_{yy} & \phi_y \\ \phi_x & \phi_y & 0 \end{vmatrix},$$

is the same determinant that we found in section 2, but now the function F also depends on the variable z. Moreover, N_2 is also the same determinant but with the partial derivatives with respect to y replaced by those with respect to z.

The sign of the quadratic form $q(k,l) := M_2\,k^2 + N_2\,l^2 + 2L_2\,kl$ is determined by the signs of M_2 and $M_2N_2 - L_2^2$. Indeed, on completing the squares, we obtain

$$q(k,l) = M_2^{-1}\Big((M_2\,k + L_2\,l)^2 + (M_2\,N_2 - L_2^2)\,l^2\Big). \tag{3.3}$$

Thus, $q(k,l)$ has a definite sign when $M_2N_2 - L_2^2 > 0$. For this to happen M_2 and N_2 must have the same sign.

When $M_2 = 0$,

$$q(k,l) = N_2^{-1}\Big((N_2\,l + L_2\,k)^2 - (L_2\,k)^2\Big),$$

or $q(k,l) = 2L_2\,kl$ if N_2 also vanishes. Clearly, $q(k,l)$ now takes both positive and negative values whenever $M_2N_2 - L_2^2 < 0$.

Now, a simple computation gives

$$M_2N_2 - L_2^2 = -\left(\phi_x\right)^2 M_3, \tag{3.4}$$

where M_3 is again the *Hessian determinant*

$$M_3 = \begin{vmatrix} F_{xx} & F_{xy} & F_{xz} & \phi_x \\ F_{xy} & F_{yy} & F_{yz} & \phi_y \\ F_{xz} & F_{yz} & F_{zz} & \phi_z \\ \phi_x & \phi_y & \phi_z & 0 \end{vmatrix}.$$

The equivalent of (2.7) thus takes the form

$$HF(x_0,y_0,z_0)(h,k,l) = -\frac{(M_2\,l + L_2\,k)^2}{\left(\phi_x\right)^2 M_2} + \frac{M_3}{M_2}\,l^2 \tag{3.5}$$

It is therefore clear that, in the problem under consideration, Lagrange's criterion is equivalent to a criterion in terms of the bordered determinants M_2 and M_3. The precise reformulation is the content of the next theorem, in which we also prove, using the simple computations leading to (3.5), that Lagrange's criterion does hold for this particular case too.

Similar considerations apply when $\phi_x = 0$, by changing the role of the variables. It is important to keep in mind that the Hessian determines the sign of the left hand side of the equivalent of (2.4) only when the quadratic form q is definite, otherwise, as in the problems without constraints, higher derivatives must be considered. For a complete discussion on this point, see the classic book by Hancock [7].

Theorem 3.1. *Assume that $\vec{x}_0$ is a critical point of the associated Lagrange function of $w = f(x,y,z)$ subject to the constraint $\phi(x,y,z) = 0$, and that $\nabla\phi(\vec{x}_0) \neq 0$. Then*

i-) *if $M_2 < 0$ and $M_3 < 0$, f attains a minimum at $\vec{x}_0$,*

ii-) *if $M_2 > 0$ and $M_3 < 0$, f attains a maximum at $\vec{x}_0$,*

iii-) *if $M_3 > 0$, f does not attain either a maximum or a minimum at $\vec{x}_0$,*

iv-) *if $M_3 = 0$, further analysis is necessary to determine the nature of $\vec{x}_0$.*

Proof. The idea is to deduce this theorem from the usual criteria in which there is no constraint for a function of two variables. We may assume that $\phi_x(\vec{x}_0) \neq 0$. By the implicit function theorem, there exists a function $x = g(y,z)$ satisfying $\phi\big(g(y,z),y,z\big) = 0$ in a neighborhood of the point (y_0, z_0), then

$$0 = \phi_x g_y + \phi_y \quad \text{and} \quad 0 = \phi_x g_z + \phi_z\,, \quad \text{so} \quad g_y = -\frac{\phi_y}{\phi_x} \quad \text{and} \quad g_z = -\frac{\phi_z}{\phi_x}\,. \tag{3.6}$$

Moreover,

$$\begin{aligned} g_{yy} &= (g_y)_y = \left[2\phi_{yx}\phi_y\phi_x - \phi_{yy}\left(\phi_x\right)^2 - \phi_{xx}\left(\phi_y\right)^2\right]\left(\phi_x\right)^{-3}, \\ g_{zz} &= (g_z)_z = \left[2\phi_{zx}\phi_z\phi_x - \phi_{zz}\left(\phi_x\right)^2 - \phi_{xx}\left(\phi_z\right)^2\right]\left(\phi_x\right)^{-3}, \\ g_{zy} &= (g_y)_z = \left[\phi_{yx}\phi_z\phi_x - \phi_{zy}\left(\phi_x\right)^2 - \phi_{xx}\phi_y\phi_z + \phi_{zx}\phi_y\phi_x\right]\left(\phi_x\right)^{-3}. \end{aligned} \tag{3.7}$$

It is clear that $\vec{x}_0$ is a critical point for the problem if (y_0, z_0) is a critical point for the function $w(y,z) = f\big(g(y,z),y,z\big)$. Thus, according to the theory of extrema without constraints, we must compute its second order partial derivatives. At first order, we get

$$w_y = f_x g_y + f_y \quad \text{and} \quad w_z = f_x g_z + f_z.$$

Hence

$$\begin{aligned} A &:= w_{yy}(y_0,z_0) = f_{xx}\left(g_y\right)^2 + 2f_{yx}g_y + f_x g_{yy} + f_{yy}, \\ B &:= w_{zy}(y_0,z_0) = f_{xx}g_y g_z + f_{zx}g_y + f_x g_{zy} + f_{yx}g_z + f_{zy}, \\ C &:= w_{zz}(y_0,z_0) = f_{xx}\left(g_z\right)^2 + 2f_{zx}g_z + f_x g_{zz} + f_{zz}. \end{aligned}$$

Using (3.6) and (3.7)

$$\begin{aligned} A = \Bigg[& f_{xx}\left(\phi_y\right)^2 - \frac{f_x}{\phi_x}\phi_{xx}\left(\phi_y\right)^2 - 2\left(f_{yx}\phi_y\phi_x - \frac{f_x}{\phi_x}\phi_{yx}\phi_y\phi_x\right) \\ & + f_{yy}\left(\phi_x\right)^2 - \frac{f_x}{\phi_x}\phi_{yy}\left(\phi_x\right)^2\Bigg]\left(\phi_x\right)^{-2}. \end{aligned}$$

As $\vec{x}_0$ is a critical point of the associated Lagrange function, i.e., it satisfies the equivalent of (1.3), $f_x(\vec{x}_0) + \lambda\phi_x(\vec{x}_0) = 0$, we see that $\lambda = -f_x(\vec{x}_0)/\phi_x(\vec{x}_0)$. In particular, we find

$$\begin{aligned} A &= \left[\left(f_{xx} + \lambda\phi_{xx}\right)\left(\phi_y\right)^2 + \left(f_{yy} + \lambda\phi_{yy}\right)\left(\phi_x\right)^2 - 2\left(f_{yx} + \lambda\phi_{yx}\right)\phi_y\phi_x\right]\left(\phi_x\right)^{-2} \\ &= \left[F_{xx}\left(\phi_y\right)^2 + F_{yy}\left(\phi_x\right)^2 - 2F_{yx}\phi_y\phi_x\right]\left(\phi_x\right)^{-2} = -M_2\left(\phi_x\right)^{-2}. \end{aligned}$$

(Have a look at (3.1) and (3.2).) Identical computations give

$$\begin{aligned} C &= \left[F_{xx}\left(\phi_z\right)^2 + F_{zz}\left(\phi_x\right)^2 - 2F_{zx}\phi_z\phi_x\right]\left(\phi_x\right)^{-2} = -N_2\left(\phi_x\right)^{-2}, \\ B &= \left[F_{xx}\phi_z\phi_y + F_{zy}\left(\phi_x\right)^2 - F_{zx}\phi_y\phi_x - F_{yx}\phi_z\phi_x\right]\left(\phi_x\right)^{-2} = -L_2\left(\phi_x\right)^{-2}. \end{aligned}$$

Thus, by (3.4)

$$\Delta := AC - B^2 = \left(M_2N_2 - L_2^2\right)\left(\phi_x\right)^{-4} = -M_3\left(\phi_x\right)^{-2}.$$

As the sign of A is the opposite to that of M_2, and that of Δ is opposite to that of M_3, the theorem clearly follows from the standard second-order derivative test for the function $w(y,z) = f\big(g(y,z),y,z\big)$. □

The attentive reader would have noticed that the proofs of theorems 2.2 and 3.1 are identical.

We point out that, when $M_3 < 0$ the three principal subdeterminants: M_2, N_2 and P_2, obtained from M_3 on eliminating, respectively, the third, the second and the first row and column, all have the same sign, so we can use any of them in the criteria above. Indeed, we saw already that M_2 and N_2 have the same sign, and, had we chosen to use the condition $\nabla\phi(\vec{x}_0)\cdot(h,k,l) = 0$ to solve, say, k instead of h, we would obtain that also M_2 and P_2 have the same sign. This conclusion can alternatively be obtained from the general theory of quadratic forms. It so happens that a quadratic form is positive definite if, and only if, all principal subdeterminants are strictly positive; for a simple proof of this fact, sometimes known as Sylvester's criterion, see, for instance, [3], or any linear algebra textbook such as [4] or [8]. In [5] there is a very complete discussion of this point. In particular, based on Gaussian elimination, Sylvester's criterion is improved by showing that if all leading principal subdeterminants of a $n\times n$ matrix A (i.e., those obtained by eliminating the last r rows and columns of A for $r = 1,\ldots,n$) are positive, then all principal subdeterminants of A are also positive.

The examples at the end of the previous section can be slightly modified to provide an illustration of the lack of criteria when $M_3 = 0$. Consider, indeed, the function $f(x,y,z) = (x+y+z)^n$, where $n \geq 3$ is an integer, subject to the condition $x^2+y^2+z^2 = 3$. It is easy to check that the critical points are $(1,1,1)$, $(-1,-1,-1)$, together with all the points of the great circle $\mathcal{C}:\left\{\begin{matrix} x+y+z=0 \\ x^2+y^2+z^2=3\end{matrix}\right\}$. Furthermore, for all points on $\mathcal{C}$ the corresponding Lagrange multiplier is $\lambda = 0$, and one readily sees that M_3 vanishes at these points. When n is even, we know that

$$f(x_0,y_0,z_0) = 0 \leq f(x,y,z),$$

for any (x,y,z) and any (x_0,y_0,z_0) in $\mathcal{C}$, therefore f attains a minimum at each point of $\mathcal{C}$. On the other hand, when n is odd, $f(x,y,z) > 0$ on the open half space above the plane $x+y+z = 0$, whereas $f(x,y,z) < 0$ on the open half space below it, therefore f attains neither a maximum nor a minimum at the points of $\mathcal{C}$.

IV. Classification of extrema of functions of three variables on a curve

Let $\vec{x}_0 := (x_0,y_0,z_0)$ be a critical point of the function $w = f(x,y,z)$ on the curve (1.1), and assume that $\nabla\phi^1(\vec{x}_0)$ and $\nabla\phi^2(\vec{x}_0)$ are linearly independent. The classification of $\vec{x}_0$, depends on the sign of $HF(x_0,y_0,z_0)(h,k,l)$, subject to the linear conditions

$$\begin{cases}\nabla\phi^1(x_0,y_0,z_0)\cdot(h,k,l) = 0,\\ \nabla\phi^2(x_0,y_0,z_0)\cdot(h,k,l) = 0.\end{cases} \tag{4.1}$$

If $\frac{\partial(\phi^1,\phi^2)}{\partial(x,y)} \neq 0$, say, then, by Cramer's rule, the solution of (4.1) is

$$h = -\frac{\begin{vmatrix}\phi^1_z & \phi^1_y\\ \phi^2_z & \phi^2_y\end{vmatrix}}{\begin{vmatrix}\phi^1_x & \phi^1_y\\ \phi^2_x & \phi^2_y\end{vmatrix}}\, l = -\frac{\dfrac{\partial(\phi^1,\phi^2)}{\partial(z,y)}}{\dfrac{\partial(\phi^1,\phi^2)}{\partial(x,y)}}\, l, \quad \text{and} \quad k = -\frac{\begin{vmatrix}\phi^1_x & \phi^1_z\\ \phi^2_x & \phi^2_z\end{vmatrix}}{\begin{vmatrix}\phi^1_x & \phi^1_y\\ \phi^2_x & \phi^2_y\end{vmatrix}}\, l = -\frac{\dfrac{\partial(\phi^1,\phi^2)}{\partial(x,z)}}{\dfrac{\partial(\phi^1,\phi^2)}{\partial(x,y)}}\, l,$$

Using that $HF(\vec{x}_0)$ is homogeneous of degree two, one arrives, after a rather long but easy computation, at the fundamental result:

$$\begin{aligned} HF(\vec{x}_0)(h,k,l) &= l^2\left(\frac{\partial(\phi^1,\phi^2)}{\partial(x,y)}\right)^{-2} HF(\vec{x}_0)\left(-\frac{\partial(\phi^1,\phi^2)}{\partial(z,y)}, -\frac{\partial(\phi^1,\phi^2)}{\partial(x,z)}, \frac{\partial(\phi^1,\phi^2)}{\partial(x,y)}\right)\\ &= l^2\left(\frac{\partial(\phi^1,\phi^2)}{\partial(x,y)}\right)^{-2}\mathcal{M}, \end{aligned} \tag{4.2}$$

in which $\mathcal{M}$ is now the *Hessian determinant*

$$\mathcal{M} = \begin{vmatrix} F_{xx} & F_{xy} & F_{xz} & \phi^1_x & \phi^2_x\\ F_{xy} & F_{yy} & F_{yz} & \phi^1_y & \phi^2_y\\ F_{xz} & F_{yz} & F_{zz} & \phi^1_z & \phi^2_z\\ \phi^1_x & \phi^1_y & \phi^1_z & 0 & 0\\ \phi^2_x & \phi^2_y & \phi^2_z & 0 & 0 \end{vmatrix}.$$

In other words, the sign of the quadratic form $HF(x_0,y_0,z_0)(h,k,l)$, on those points satisfying (4.1), is the same as the sign of $\mathcal{M}$, The equivalent reformulation of Lagrange's criterion so obtained. is given in the next theorem. When $\frac{\partial(\phi^1,\phi^2)}{\partial(x,y)} = 0$ instead, one clearly obtains the same result by changing the roles of the variables, since one of the three Jacobians of ϕ^1 and ϕ^2 does not vanish.

Theorem 4.1. *Suppose that $\vec{x}_0$ is a critical point of the Lagrange function associated with $w = f(x,y,z)$ with the side conditions $\phi^1(x,y,z) = 0$ and $\phi^2(x,y,z) = 0$, and that $\nabla\phi^1(\vec{x}_0)$ and $\nabla\phi^2(\vec{x}_0)$ are linearly independent. Then*

i-) *if $\mathcal{M} > 0$, f attains a minimum at $\vec{x}_0$,*

ii-) *if $\mathcal{M} < 0$, f attains a maximum at $\vec{x}_0$,*

iii-) *when $\mathcal{M} = 0$, further considerations are required to classify the critical point $\vec{x}_0$.*

Proof. This theorem follows from the second derivative test for a function of one variable, upon using the same argument as in the proof of theorem 2.2, or theorem 3.1, and taking into consideration the key result (4.2). Thus, we focus on the third assertion of the theorem. Let $f(x,y,z) = (x-y+z)^n$, $\phi^1(x,y,z) = x^2+y^2+z^2-2$ and $\phi^2(x,y,z) = -x+y+z$. The conditions determine a great circle $\mathcal{C}$. The critical points of this particular problem are $(\frac{\sqrt{3}}{3}, -\frac{\sqrt{3}}{3}, \frac{2\sqrt{3}}{3})$, $(-\frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3}, -\frac{2\sqrt{3}}{3})$, $(1,1,0)$ and $(-1,-1,0)$. We find that $\mathcal{M}(1,1,0) = 0$ and $\mathcal{M}(-1,-1,0) = 0$. Now, f vanishes at $(1,1,0)$ and $(-1,-1,0)$, and since f is positive when

n is even, it follows that f attains minima at $(1,1,0)$ and $(-1,-1,0)$, in those cases. On the other hand, when n is odd, notice that in the half circle of $\mathcal{C}$ from $(1,1,0)$ to $(-1,-1,0)$, when one goes through $\mathcal{C}$ clockwise if regarded from the point $(-9,9,9)$, f is positive because both $z>0$ and $x-y>0$ on those points, whereas on the arc from $(-1,-1,0)$ to $(1,1,0)$, $z<0$ and $x-y<0$, so f is negative. Therefore f attains neither a maximum nor a minimum at $(1,1,0)$, and likewise at $(-1,-1,0)$. □

Naturally, theorems 2.2, 3.1 and 4.1 can be generalized to classify the critical points of functions of n variables subject to m conditions, provided $m<n$. All that is required is a careful study of the sign of the quadratic forms under the presence of homogeneous linear conditions; which turns out to depend only on the dimension, and the sign of several Hessian determinants, similar to those considered above. As far as we know, the first criterion related to Hessian determinants was given by Hancock in 1917 [7, section 89]. Hancock, however, did not write explicitly the Hessian determinants, but rather gave his criterion in terms of the signs of the coefficients of certain polynomials of a variable e:

Theorem 4.2 (Hancock). *Under the same hypothesis of theorem 2.1, consider the polynomial of degree $n-m$ given by*

$$P_{n,m}(e) := \begin{vmatrix} F_{x_1x_1}-e & F_{x_1x_2} & \cdots & F_{x_1x_n} & \phi^1_{x1} & \cdots & \phi^m_{x1} \\ F_{x_2x_1} & F_{x_2x_2}-e & \cdots & F_{x_2x_n} & \phi^1_{x2} & \cdots & \phi^m_{x2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ F_{x_nx_1} & F_{x_nx_2} & \cdots & F_{x_nx_n}-e & \phi^1_{xn} & \cdots & \phi^m_{xn} \\ \phi^1_{x1} & \phi^1_{x2} & \cdots & \phi^1_{xn} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \phi^m_{x1} & \phi^m_{x2} & \cdots & \phi^m_{xn} & 0 & \cdots & 0 \end{vmatrix}.$$

i-) *If the coefficients of $P_{n,m}(e)$ alternate signs, f attains a minimum at $\vec{x}_0$.*

ii-) *If the coefficients of $P_{n,m}(e)$ have all the same sign, f attains a maximum at $\vec{x}_0$.*

One can easily verify that

$$P_{2,1}(e) = \left[(\phi_x)^2 + (\phi_y)^2\right]\, e + M,$$
$$P_{3,1}(e) = -\left[(\phi_x)^2 + (\phi_y)^2 + (\phi_z)^2\right]\, e^2 - (M_2+N_2+P_2)\, e + M_3,$$
$$P_{3,2}(e) = -\left[\left(\frac{\partial(\phi^1,\phi^2)}{\partial(x,y)}\right)^2 + \left(\frac{\partial(\phi^1,\phi^2)}{\partial(x,z)}\right)^2 + \left(\frac{\partial(\phi^1,\phi^2)}{\partial(y,z)}\right)^2\right]\, e + \mathcal{M}.$$

Thus, a careful bookkeeping shows that Hancock's criterion and the criterion in terms of the Hessian determinants are equivalent, at least for the cases considered above; the main difference being that Hancock emphasized the symmetry of the variables, as each plays the same role in all coefficients of the polynomials $P_{n,m}$. For instance, bearing in mind that M_2, N_2 and P_2 have the same sign when the quadratic form q in (3.3) (or one of its equivalent forms) is definite, we see that Hancock's criterion entails that f attains a minimum at

$\vec{x}_0$ when $M_2 < 0$ and $M_3 < 0$, and a maximum if $M_2 > 0$ and $M_3 < 0$, in agreement with theorem 3.1. It is interesting to point out that Caratheodory, in 1935 [3], gave a rigorous proof of Lagrange's criterion using the polynomials $P_{n,m}$, but did not write an explicit criterion, such as Hancock's or the one in terms of bordered determinants.

There are several ways to implement Lagrange's criterion directly. When there are m constraints, the linear independence of the vectors $\nabla\phi^1(\vec{x}_0), \ldots, \nabla\phi^m(\vec{x}_0)$ means that the solution space of (2.2) has dimension $n - m$. If $\vec{v}_1, \ldots, \vec{v}_{n-m}$ is a basis of such vector space, and

$$N = \begin{pmatrix} F_{x_1x_1} & \cdots & F_{x_1x_n} \\ \vdots & \ddots & \vdots \\ F_{x_nx_1} & \cdots & F_{x_nx_n} \end{pmatrix},$$

so that $HF(\vec{x}_0)(\vec{h}) = \vec{h} \cdot N\vec{h}$, then

$$HF(\vec{x}_0)(h_1\vec{v}_1 + \cdots + h_{n-m}\vec{v}_{n-m}) = \sum_{i,j=1}^{n-m} \vec{v}_i \cdot N\vec{v}_j \, h_i h_j \,,$$

in other words, $HF(\vec{x}_0)(h_1\vec{v}_1 + \cdots + h_{n-m}\vec{v}_{n-m})$ may be regarded as the quadratic form in $\mathbb{R}^{n-m}$ associated with the symmetric matrix

$$K = \begin{pmatrix} \vec{v}_1 \cdot N\vec{v}_1 & \vec{v}_1 \cdot N\vec{v}_2 & \cdots & \vec{v}_1 \cdot N\vec{v}_{n-m} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{v}_{n-m} \cdot N\vec{v}_1 & \vec{v}_{n-m} \cdot N\vec{v}_2 & \cdots & \vec{v}_{n-m} \cdot N\vec{v}_{n-m} \end{pmatrix}.$$

There are at least two possible ways to determine the sign of this new quadratic form: one can use Sylvester criteria in terms of the principal minors of K, see [4] or [5], or one can look at the signs of the eigenvalues of K [1].

However, quite often in practice, especially in lower dimensions, it is easier to use the system (2.2) to solve $n - m$ of the h_i in terms of the remaining h_j; to plug these relations into $HF(\vec{x}_0)(\vec{h})$, which amounts to obtaining the quadratic form associated to K; and to perform simple algebraic manipulations on this new quadratic form, such as completing the squares, to analyze directly the sign of this quadratic form, much in the spirit of reducing a quadratic form to its canonical form, which is what is behind all these criteria and, in a way, was Lagrange's original idea.

Acknowledgments

We thank Joseph C. Várilly for useful comments and the Vicerrectoría de Investigación de la Universidad de Costa Rica for support.

V. References

[1] T. Apostol, *Mathematical Analysis: A Modern Approach to Advanced Calculus*, Addison-Wesley, Reading, MA, 6th printing, 1973.

[2] R. G. Bartle, *The Elements of Real Analysis*, John Wiley and Sons, New York, 2nd edition, 1976.

[3] C. Caratheodory, *Variationsrechnung und Partielle Differentialgleichungen erster Ordnung*, B. G. Teubner, Berlin, 1935. English version: *Calculus of Variations and Partial Differential Equations of the First Order*, 3rd edition, Chelsea, New York, 1989.

[4] W. L. Ferrar, *Algebra: a Textbook of Determinants, Matrices and Algebraic Forms*, 2nd edition, Oxford University Press, New York, 1957.

[5] F. R. Gantmacher, *The Theory of Matrices*, Vol. I, Chelsea, New York, 1959.

[6] E. Goursat, *A Course in Mathematical Analysis*, Vol. I, Dover Books, New York, 1959.

[7] H. Hancock, *Theory of Maxima and Minima*, Dover Books, New York, 1960. Originally published by Ginn and Company, 1917.

[8] K. Hoffman and R. Kunze, *Linear Algebra*, 2nd edition, Prentice-Hall, Englewood Cliffs, NJ, 1971.

[9] J. L. Lagrange, *Théorie des Fonctions Analytiques*, J. de l'École Polytechnique, neuvième cahier, tome III, 1797.