

# Lagrange Multipliers

Com S 477/577

Nov 18, 2008

## 1 Introduction

We turn now to the study of minimization with constraints. More specifically, we will tackle the following problem:

$$\begin{array}{ll}\text{minimize} & f(\mathbf{x}) \\ \text{subject to} & h_1(\mathbf{x}) = 0 \\ & \vdots \\ & h_m(\mathbf{x}) = 0\end{array}$$

where  $\mathbf{x} \in \Omega \subset \mathbb{R}^n$ , and the functions  $f, h_1, \dots, h_m$  are continuous, and usually assumed to be in  $C^2$  (i.e., with continuous second partial derivatives).

When  $f$  and  $h_j$ 's are linear, the problem is a linear programming one and can be solved using the simplex algorithm. Hence we would like to focus on the case that these functions are nonlinear.

In order to gain some intuition, let us look at the case where  $n = 2$  and  $m = 1$ . The problem becomes

$$\begin{array}{ll}\text{minimize} & f(x, y) \\ \text{subject to} & h(x, y) = 0, \quad x, y \in \mathbb{R}\end{array}$$

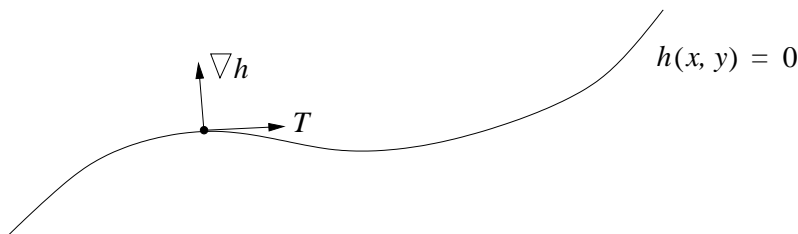
The constraint  $h(x, y) = 0$  defines a curve as shown below. Differentiate the equation with respect to  $x$ :

$$\frac{\partial h}{\partial x} + \frac{\partial h}{\partial y} \frac{dy}{dx} = 0.$$

The tangent of the curve is  $T(x, y) = (1, \frac{dy}{dx})$ . And the gradient of the curve is  $\nabla h = (\frac{\partial h}{\partial x}, \frac{\partial h}{\partial y})$ . So the above equation states that

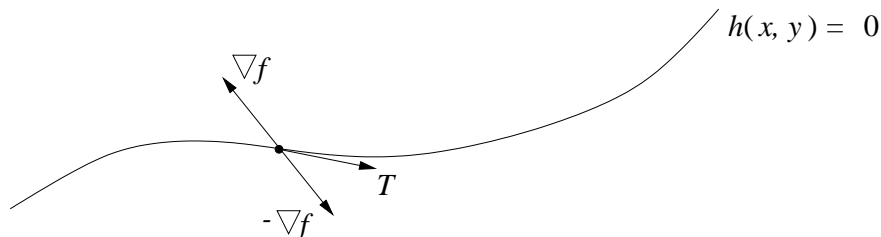
$$T \cdot \nabla h = 0;$$

namely, the tangent of the curve must be normal to the gradient all the time. Suppose we are at a point on the curve. To stay on the curve, any motion must be along the tangent  $T$ .



In order to increase or decrease  $f(x, y)$ , motion along the constraint curve must have a component along the gradient of  $f$ , that is,

$$\nabla f \cdot T \neq 0.$$

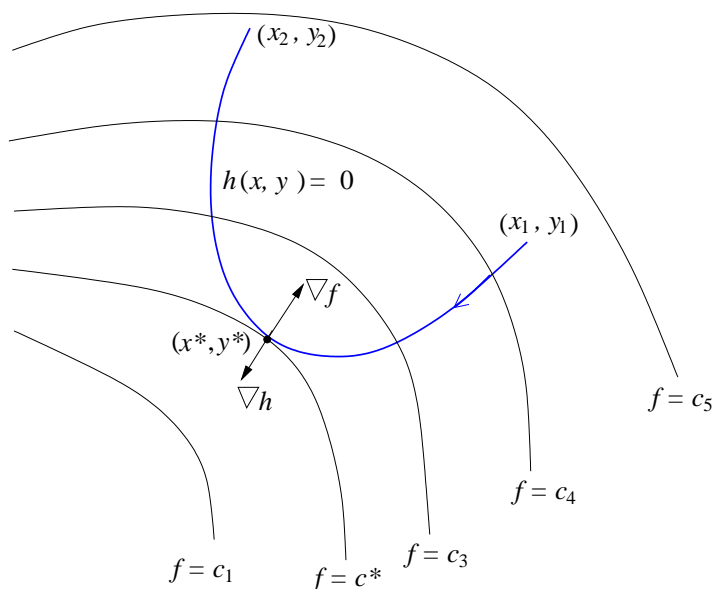


At an extremum of  $f$ , a differential motion should not yield a component of motion along  $\nabla f$ . Thus  $T$  is orthogonal to  $\nabla f$ ; in other words, the condition

$$\nabla f \cdot T = 0$$

must hold. Now  $T$  is orthogonal to both gradients  $\nabla f$  and  $\nabla h$  at an extrema. This means that  $\nabla f$  and  $\nabla h$  must be parallel. Phrased differently, there exists some  $\lambda \in \mathbb{R}$  such that

$$\nabla f + \lambda \nabla h = 0. \quad (1)$$



The figure above explains condition (1) by superposing the curve  $h(x, y) = 0$  onto the family of *level curves* of  $f(x, y)$ , that is, the collection of curves  $f(x, y) = c$ , where  $c$  is any real number in the range of  $f$ . In the figure,  $c_5 > c_4 > c_3 > c^* > c_1$ . The tangent of a level curve is always orthogonal to the gradient  $\nabla f$ . Otherwise moving along the curve would result in an increase or decrease of the value of  $f$ . Imagine a point moving on the curve  $h(x, y) = 0$  from  $(x_1, y_1)$  to  $(x_2, y_2)$ . Initially, the motion has a component along the negative gradient direction  $-\nabla f$ , resulting in the decrease

of the value of  $f$ . This component becomes smaller and smaller. When the moving point reaches  $(x^*, y^*)$ , the motion is orthogonal to the gradient. From that point on, the motion starts having a component along the gradient direction  $\nabla f$  so the value of  $f$  increases. Thus at  $(x^*, y^*)$ ,  $f$  achieves its local minimum. The motion is in the tangential direction of the curve  $h(x, y) = 0$ , which is orthogonal to the gradient  $\nabla h$ . Therefore at the point  $(x^*, y^*)$  the two gradients  $\nabla f$  and  $\nabla h$  must be collinear. This is what equation (1) says. Let  $c^*$  be the local minimum achieved at  $(x^*, y^*)$ . It is clear that the two curves  $f(x, y) = c^*$  and  $h(x, y) = 0$  are tangent at  $(x^*, y^*)$ .

Suppose we find the set  $S$  of points satisfying the equations

$$\begin{aligned} h(x, y) &= 0, \\ \nabla f + \lambda \nabla h &= 0, \quad \text{for some } \lambda. \end{aligned}$$

Then  $S$  contains the extremal points of  $f$  subject to the constraints  $h(x, y) = 0$ . The above two equations constitute a nonlinear system in the variables  $x, y, \lambda$ . It can be solved using numerical techniques, for example, Newton's method.

## 2 Lagrangian

It is convenient to introduce the *Lagrangian* associated with the constrained problem, defined as

$$F(x, y, \lambda) = f(x, y) + \lambda h(x, y).$$

Note that

$$\nabla F = \begin{pmatrix} \frac{\partial f}{\partial x} + \lambda \frac{\partial h}{\partial x} \\ \frac{\partial f}{\partial y} + \lambda \frac{\partial h}{\partial y} \\ h \end{pmatrix}^T = (\nabla f + \lambda \nabla h, h).$$

Thus setting  $\nabla F = 0$  yields the same system of nonlinear equations we derived earlier.

The value  $\lambda$  is known as the *Lagrange multiplier*. The approach of constructing the Lagrangians and setting its gradient to zero is known as the method of Lagrange multipliers.

EXAMPLE 1 Find the extremal values of the function  $f(x, y) = xy$  subject to the constraint

$$h(x, y) = \frac{x^2}{8} + \frac{y^2}{2} - 1 = 0.$$

We first construct the Lagrangian and find its gradient:

$$\begin{aligned} F(x, y, \lambda) &= xy + \lambda \left( \frac{x^2}{8} + \frac{y^2}{2} - 1 \right), \\ \nabla F(x, y, \lambda) &= \begin{pmatrix} y + \frac{\lambda x}{4} \\ x + \lambda y \\ \frac{x^2}{8} + \frac{y^2}{2} - 1 \end{pmatrix} = \mathbf{0}. \end{aligned}$$

The above leads to three equations

$$y + \frac{\lambda x}{4} = 0, \tag{2}$$

$$x + \lambda y = 0, \tag{3}$$

$$x^2 + 4y^2 = 8. \tag{4}$$

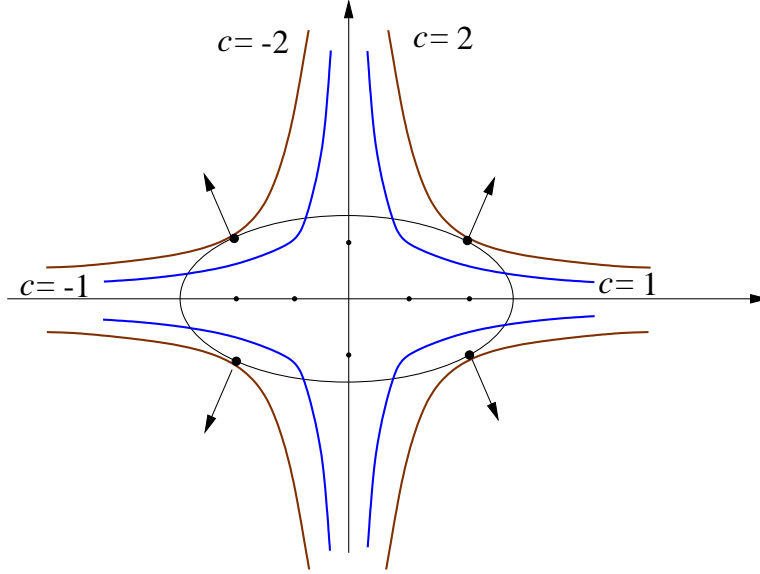
Combining (2) and (3) yields

$$\lambda^2 = 4 \quad \text{and} \quad \lambda = \pm 2.$$

Thus  $x = \pm 2y$ . Substituting this equation into (4) gives us

$$y = \pm 1 \quad \text{and} \quad x = \pm 2.$$

So there are four extremal points of  $f$  subject to the constraint  $h$ :  $(2, 1)$ ,  $(-2, -1)$ ,  $(2, -1)$ , and  $(-2, 1)$ . The maximum value 2 is achieved at the first two points while the minimum value  $-2$  is achieved at the last two points.



Graphically, the constraint  $h$  defines an ellipse. The constraint contours of  $f$  are the hyperbolas  $xy = c$ , with  $|c|$  increasing as the curves move out from the origin.

### 3 General Formulation

Now we would like to generalize to the case with multiple constraints. Let  $\mathbf{h} = (h_1, \dots, h_m)^T$  be a function from  $\mathbb{R}^n$  to  $\mathbb{R}^m$ . Consider the constrained optimization problem below.

$$\begin{aligned} & \text{minimize} && f(\mathbf{x}) \\ & \text{subject to} && \mathbf{h}(\mathbf{x}) = \mathbf{0}. \end{aligned}$$

Each constraint equation  $h_j(\mathbf{x}) = 0$  defines a constraint *hypersurface*  $S$  in the space  $\mathbb{R}^n$ . And this surface is smooth provided  $h_j(\mathbf{x}) \in C^1$ .

A *curve* on  $S$  is a family of points  $\mathbf{x}(t) \in S$  with  $a \leq t \leq b$ . the curve is differentiable if  $\frac{d\mathbf{x}}{dt}$  exists, and twice differentiable if  $\frac{d^2\mathbf{x}}{dt^2}$  exists. The curve passes through a point  $\mathbf{x}^*$  if  $\mathbf{x}^* = \mathbf{x}(t^*)$  for some  $t^*$ ,  $a \leq t^* \leq b$ .

The *tangent space* at  $\mathbf{x}^*$  is the subspace of  $\mathbb{R}^n$  spanned by the tangents  $\frac{d\mathbf{x}}{dt}(t^*)$  of all curves  $\mathbf{x}(t)$  on  $S$  such that  $\mathbf{x}(t^*) = \mathbf{x}^*$ . In other words, the tangent space is the set of the derivatives at  $\mathbf{x}^*$  of all surface curves through  $\mathbf{x}^*$ . Denote this subspace as  $\mathcal{T}$ .

A point  $\mathbf{x}$  satisfying  $\mathbf{h}(\mathbf{x}) = \mathbf{0}$  is a *regular point* of the constraint if the gradient vectors  $\nabla h_1(\mathbf{x}), \dots, \nabla h_m(\mathbf{x})$  are linearly independent.

From our previous intuition, we expect that  $\nabla f \cdot \mathbf{v} = 0$  for all  $\mathbf{v} \in \mathcal{T}$  at an extremum. This implies that  $\nabla f$  lies in the orthogonal complement  $\mathcal{T}^\perp$  of  $\mathcal{T}$ . We would like to claim that  $\nabla f$  can be composed from a linear combination of the  $\nabla h_i$ 's. this is only valid provided that these gradients span  $\mathcal{T}^\perp$ , which is true when the extremal point is regular.

**Theorem 1** *At a regular point  $\mathbf{x}$  of the surface  $S$  defined by  $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ , the tangent space is the same as*

$$\{\mathbf{y} \mid \nabla \mathbf{h}(\mathbf{x})\mathbf{y} = \mathbf{0}\},$$

where the matrix

$$\nabla \mathbf{h} = \begin{pmatrix} \nabla h_1 \\ \vdots \\ \nabla h_m \end{pmatrix}.$$

The rows of the matrix  $\nabla \mathbf{h}(\mathbf{x})$  are the gradient vectors  $\nabla h_j(\mathbf{x})$ ,  $j = 1, \dots, m$ . The theorem says that the tangent space at  $\mathbf{x}$  is equal to the nullspace of  $\nabla \mathbf{h}(\mathbf{x})$ . Thus its orthogonal complement  $\mathcal{T}^\perp$  must equal the row space of  $\nabla \mathbf{h}(\mathbf{x})$ . Hence the vectors  $\nabla h_j(\mathbf{x})$  span  $\mathcal{T}^\perp$ .

EXAMPLE 2. Suppose  $h(x_1, x_2) = x_1$ . Then  $h(\mathbf{x}) = 0$  yields the  $x_2$  axis. And  $\nabla h = (1, 0)$  at all points. So every  $\mathbf{x} \in \mathbb{R}^2$  is regular. The tangent space is also the  $x_2$  axis and has dimension 1.

If instead  $h(x_1, x_2) = x_1^2$  then  $h(\mathbf{x}) = 0$  still defines the  $x_2$  axis. On this axis,  $\nabla h = (2x_1, 0) = (0, 0)$ . Thus no point is regular. The dimension of  $\mathcal{T}$ , which is the  $x_2$  axis, is still one, but the dimension of the space  $\{\mathbf{y} \mid \nabla h \cdot \mathbf{y} = 0\}$  is two.

**Lemma 2** *Let  $\mathbf{x}^*$  be a local extremum of  $f$  subject to the constraints  $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ . Then for all  $\mathbf{y}$  in the tangent space of the constraint surface at  $\mathbf{x}^*$ ,*

$$\nabla f(\mathbf{x}^*)\mathbf{y} = 0.$$

The next theorem states that the Lagrange multiplier method as a necessary condition on an extremum point.

**Theorem 3 (First-Order Necessary Conditions)** *Let  $\mathbf{x}^*$  be a local extremum point of  $f$  subject to the constraints  $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ . Assume further that  $\mathbf{x}^*$  is a regular point of these constraints. Then there is a  $\boldsymbol{\lambda} \in \mathbb{R}^n$  such that*

$$\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0}.$$

The first order necessary conditions together with the constraints

$$\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$$

give a total of  $n + m$  equations in  $n + m$  variables  $\mathbf{x}^*$  and  $\boldsymbol{\lambda}$ . Thus a unique solution can be determined at least locally.

EXAMPLE 3. We seek to construct a cardboard box of maximum volume, given a fixed area of cardboard.

Denoting the dimensions of the box by  $x, y, z$ , the problem can be expressed as

$$\begin{aligned} & \text{maximize} && xyz \\ & \text{subject to} && xy + yz + xz = \frac{c}{2}, \end{aligned}$$

where  $c > 0$  is the given area of cardboard. Consider the Lagrangian  $xyz + \lambda(xy + yz + xz - \frac{c}{2})$ . The first-order necessary conditions are easily found to be

$$yz + \lambda(y + z) = 0, \tag{5}$$

$$xz + \lambda(x + z) = 0, \tag{6}$$

$$xy + \lambda(x + y) = 0, \tag{7}$$

together with the original constraint. Before solving the above equations, we note that their sum is

$$(xy + yz + xz) + 2\lambda(x + y + z) = 0,$$

which, given the constraint, becomes

$$c/2 + 2\lambda(x + y + z) = 0.$$

Hence it is clear that  $\lambda \neq 0$ . Neither of  $x, y, z$  can be zero since if either is zero, all must be so according to (5)–(7).

To solve the equations (5)–(7), multiply (5) by  $x$  and (6) by  $y$ , and then subtract the two to obtain

$$\lambda(x - y)z = 0.$$

Operate similarly on the second and the third to obtain

$$\lambda(y - z)x = 0.$$

Since no variables can be zero, it follows that

$$x = y = z = \sqrt{\frac{c}{6}}$$

is the unique solution to the necessary conditions. The box must be a cube.

We can derive second-order conditions for constrained problems, assuming  $f$  and  $\mathbf{h}$  are twice continuously differentiable.

**Theorem 4 (Second-Order Necessary Conditions)** *Suppose that  $\mathbf{x}^*$  is a local minimum of  $f$  subject to  $\mathbf{h}(\mathbf{x}) = \mathbf{0}$  and that  $\mathbf{x}^*$  is a regular point of these constraints. Then there is a  $\boldsymbol{\lambda} \in \mathbb{R}^m$  such that*

$$\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0}.$$

*The matrix*

$$L(\mathbf{x}^*) = Hf(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i Hh_i(\mathbf{x}^*) \tag{8}$$

*is positive semidefinite on the tangent space  $\{\mathbf{y} \mid \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}\}$ .*

**Theorem 5 (Second-Order Sufficient Conditions)** *Suppose there is a point  $\mathbf{x}^*$  satisfying  $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$ , and a  $\boldsymbol{\lambda}$  such that*

$$\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0}.$$

*Suppose also that the matrix  $L(\mathbf{x}^*)$  defined in (8) is positive definite on the tangent space  $\{\mathbf{y} \mid \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}\}$ . Then  $\mathbf{x}^*$  is a strict local minimum of  $f$  subject to  $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ .*

EXAMPLE 4. consider the problem

$$\begin{aligned} & \text{minimize} && x_1x_2 + x_2x_3 + x_1x_3 \\ & \text{subject to} && x_1 + x_2 + x_3 = 3 \end{aligned}$$

The first order necessary conditions become

$$\begin{aligned} x_2 + x_3 + \lambda &= 0, \\ x_1 + x_3 + \lambda &= 0, \\ x_1 + x_2 + \lambda &= 0. \end{aligned}$$

We can solve these three equations together with the one constraint equation and obtain

$$x_1 = x_2 = x_3 = 1 \quad \text{and} \quad \lambda = -2.$$

Thus  $\mathbf{x}^* = (1, 1, 1)^T$ .

Now we need to resort to the second-order sufficient conditions to determine if the problem achieves a local maximum or minimum at  $x_1 = x_2 = x_3 = 1$ . We find the matrix

$$\begin{aligned} L(\mathbf{x}^*) &= Hf(\mathbf{x}^*) + \lambda Hh(\mathbf{x}^*) \\ &= \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \end{aligned}$$

is neither positive nor negative definite. On the tangent space  $M = \{ \mathbf{y} \mid y_1 + y_2 + y_3 = 0 \}$ , however, we note that

$$\begin{aligned} \mathbf{y}^T L \mathbf{y} &= y_1(y_2 + y_3) + y_2(y_1 + y_3) + y_3(y_1 + y_2) \\ &= -(y_1^2 + y_2^2 + y_3^2) \\ &< 0, \quad \text{for all } \mathbf{y} \neq \mathbf{0}. \end{aligned}$$

Thus  $L$  is negative definite on  $M$  and the solution 3 we found is at least a local maximum.

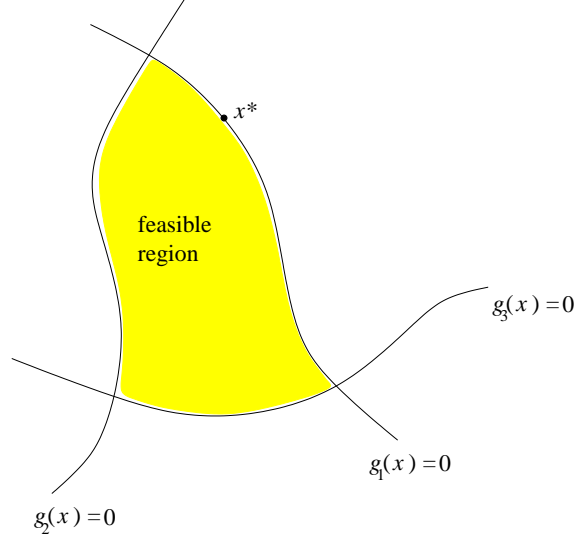
## 4 Inequality Constraints

Finally, we address problems of the general form

$$\begin{aligned} & \text{minimize} && f(\mathbf{x}) \\ & \text{subject to} && \mathbf{h}(\mathbf{x}) = \mathbf{0} \\ & && \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \end{aligned}$$

where  $\mathbf{h} = (h_1, \dots, h_m)^T$  and  $\mathbf{g} = (g_1, \dots, g_p)^T$ .

A fundamental concept that provides a great deal of insight as well as simplifies the required theoretical development is that of an *active constraint*. An inequality constraint  $g_i(\mathbf{x}) \leq 0$  is said to be *active* at a feasible point  $\mathbf{x}$  if  $g_i(\mathbf{x}) = 0$  and *inactive* at  $\mathbf{x}$  if  $g_i(\mathbf{x}) < 0$ . By convention we refer to any equality constraint  $h_i(\mathbf{x}) = 0$  as active at any feasible point. The constraints active at a feasible point  $\mathbf{x}$  restrict the domain of feasibility in neighborhoods of  $\mathbf{x}$ . Therefore, in studying the properties of a local minimum point, it is clear that attention can be restricted to the active constraints. This is illustrated in the figure below where local properties satisfied by the solution  $\mathbf{x}^*$  obviously do not depend on the inactive constraints  $g_2$  and  $g_3$ .



Assume that the functions  $f$ ,  $\mathbf{h} = (h_1, \dots, h_m)^T$ ,  $\mathbf{g} = (g_1, \dots, g_p)^T$  are twice continuously differentiable. Let  $\mathbf{x}^*$  be a point satisfying the constraints

$$\mathbf{h}(\mathbf{x}^*) = \mathbf{0} \quad \text{and} \quad \mathbf{g}(\mathbf{x}^*) \leq \mathbf{0},$$

and let  $J = \{j \mid g_j(\mathbf{x}^*) = 0\}$ . Then  $\mathbf{x}^*$  is said to be a *regular point* of the above constraints if the gradient vectors  $\nabla h_i(\mathbf{x}^*)$ ,  $\nabla g_j(\mathbf{x}^*)$ ,  $1 \leq i \leq m$ ,  $j \in J$  are linearly independent. Now suppose this regular point  $\mathbf{x}^*$  is also a relative minimum point for the original problem (1). Then it is shown that there exists a vector  $\boldsymbol{\lambda} \in \mathbb{R}^m$  and a vector  $\boldsymbol{\mu} \in \mathbb{R}^p$  with  $\boldsymbol{\mu} \geq \mathbf{0}$  such that

$$\begin{aligned} \nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) + \boldsymbol{\mu}^T \nabla \mathbf{g}(\mathbf{x}^*) &= \mathbf{0}; \\ \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}^*) &= 0. \end{aligned}$$

Since  $\boldsymbol{\mu} \geq \mathbf{0}$  and  $\mathbf{g}(\mathbf{x}^*) \leq \mathbf{0}$ , the second constraint above is equivalent to the statement that a component of  $\boldsymbol{\mu}$  may be nonzero only if the corresponding constraint is active.

To find a solution, we enumerate various combinations of *active* constraints, that is, constraints where equalities are attained at  $\mathbf{x}^*$ , and check the signs of the resulting Lagrange multipliers.

There are a number of distinct theories concerning this problem, based on various regularity conditions or constraint qualifications, which are directed toward obtaining definitive general statements of necessary and sufficient conditions. One can by no means pretend that all such results can be obtained as minor extensions of the theory for problems having equality constraints only. To date, however, their use has been limited to small-scale programming problems of two or three variables. We refer you to [1] for more on the subject.

## References

- [1] D. G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley, 2nd edition, 1984.
- [2] T. Abell. Lecture notes. The Robotics Institute, School of Computer Science, Carnegie Mellon University, 1992.