Optimization using Calculus

Optimization of Functions of Multiple Variables subject to Equality Constraints

D Nagesh Kumar, IISc

Objectives

- Optimization of functions of multiple variables subjected to equality constraints using
  - the method of constrained variation, and
  - the method of Lagrange multipliers.

Constrained optimization

A function of multiple variables, $f(x)$, is to be optimized subject to one or more equality constraints of many variables. These equality constraints, $g_j(x)$, may or may not be linear. The problem statement is as follows:

Maximize (or minimize) $f(\mathbf{X})$, subject to $g_j(\mathbf{X}) = 0, \; j = 1, 2, \ldots, m$

where

$$\mathbf{X} = \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Bmatrix}$$

Constrained optimization (contd.)

- With the condition that $m \leq n$; or else if $m > n$ then the problem becomes an over defined one and there will be no solution. Of the many available methods, the method of constrained variation and the method of using Lagrange multipliers are discussed.

Solution by method of Constrained Variation

- For the optimization problem defined above, let us consider a specific case with $n = 2$ and $m = 1$ before we proceed to find the necessary and sufficient conditions for a general problem using Lagrange multipliers. The problem statement is as follows:

Minimize $f(x_1,x_2)$, subject to $g(x_1,x_2) = 0$

- For $f(x_1,x_2)$ to have a minimum at a point $X^* = [x_1^*,x_2^*]$, a necessary condition is that the total derivative of $f(x_1,x_2)$ must be zero at $[x_1^*,x_2^*]$.

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 = 0 \qquad (1)$$

Method of Constrained Variation (contd.)

- Since $g(x_1^*, x_2^*) = 0$ at the minimum point, variations dx_1 and dx_2 about the point $[x_1^*, x_2^*]$ must be *admissible variations*, i.e. the point lies on the constraint:

$$g(x_1^* + dx_1 , x_2^* + dx_2) = 0 \qquad (2)$$

assuming dx_1 and dx_2 are small the Taylor series expansion of this gives us

$$g(x_1^* + dx_1 , x_2^* + dx_2)$$

$$= g(x_1^*, x_2^*) + \frac{\partial g}{\partial x_1}(x_1^*, x_2^*)\, dx_1 + \frac{\partial g}{\partial x_2}(x_1^*, x_2^*)\, dx_2 = 0$$

(3)

Method of Constrained Variation (contd.)

or $$dg = \frac{\partial g}{\partial x_1} dx_1 + \frac{\partial g}{\partial x_2} dx_2 = 0 \text{ at } [x_1^*, x_2^*] \quad (4)$$

which is the condition that must be satisfied for all *admissible variations*.

Assuming $\partial g / \partial x_2 \neq 0$, (4) can be rewritten as

$$dx_2 = -\frac{\partial g / \partial x_1}{\partial g / \partial x_2}(x_1^*, x_2^*) dx_1 \quad (5)$$

Method of Constrained Variation (contd.)

(5) indicates that once variation along x_1 (dx_1) is chosen arbitrarily, the variation along x_2 (dx_2) is decided automatically to satisfy the condition for the *admissible variation.* Substituting equation (5) in (1) we have:

$$df = \left(\frac{\partial f}{\partial x_1} - \frac{\partial g / \partial x_1}{\partial g / \partial x_2} \frac{\partial f}{\partial x_2} \right) \Bigg|_{(x_1^*,\, x_2^*)} dx_1 = 0 \qquad (6)$$

The equation on the left hand side is called the constrained variation of *f*. Equation (5) has to be satisfied for all dx_1, hence we have

$$\left(\frac{\partial f}{\partial x_1} \frac{\partial g}{\partial x_2} - \frac{\partial f}{\partial x_2} \frac{\partial g}{\partial x_1} \right) \Bigg|_{(x_1^*,\, x_2^*)} = 0 \qquad (7)$$

This gives us the necessary condition to have [x_1^*, x_2^*] as an extreme point (maximum or minimum)

Solution by method of Lagrange multipliers

Continuing with the same specific case of the optimization problem with n = 2 and m = 1 we define a quantity λ, called the *Lagrange multiplier* as

$$\lambda = -\left.\frac{\partial f / \partial x_2}{\partial g / \partial x_2}\right|_{(x_1^*,\, x_2^*)} \tag{8}$$

Using this in (5)

$$\left.\left(\frac{\partial f}{\partial x_1} + \lambda \frac{\partial g}{\partial x_1}\right)\right|_{(x_1^*,\, x_2^*)} = 0 \tag{9}$$

And (8) written as

$$\left.\left(\frac{\partial f}{\partial x_2} + \lambda \frac{\partial g}{\partial x_2}\right)\right|_{(x_1^*,\, x_2^*)} = 0 \tag{10}$$

Solution by method of Lagrange multipliers…contd.

Also, the constraint equation has to be satisfied at the extreme point

$$g(x_1, x_2)\big|_{(x_1^*, x_2^*)} = 0 \qquad (11)$$

Hence equations (9) to (11) represent the necessary conditions for the point [x_1*, x_2*] to be an extreme point.

Note that λ could be expressed in terms of $\partial g / \partial x_1$ as well and $\partial g / \partial x_1$ has to be non-zero.

Thus, these necessary conditions require that at least one of the partial derivatives of $g(x_1 , x_2)$ be non-zero at an extreme point.

Solution by method of Lagrange multipliers…contd.

The conditions given by equations (9) to (11) can also be generated by constructing a functions **L,** known as the Lagrangian function, as

$$L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda g(x_1, x_2) \tag{12}$$

Alternatively, treating **L** as a function of x_1,x_2 and λ, the necessary conditions for its extremum are given by

$$\begin{aligned}
\frac{\partial L}{\partial x_1}(x_1, x_2, \lambda) &= \frac{\partial f}{\partial x_1}(x_1, x_2) + \lambda \frac{\partial g}{\partial x_1}(x_1, x_2) = 0 \\
\frac{\partial L}{\partial x_2}(x_1, x_2, \lambda) &= \frac{\partial f}{\partial x_2}(x_1, x_2) + \lambda \frac{\partial g}{\partial x_2}(x_1, x_2) = 0 \qquad (13) \\
\frac{\partial L}{\partial \lambda}(x_1, x_2, \lambda) &= g(x_1, x_2) = 0
\end{aligned}$$

Necessary conditions for a general problem

For a general problem with *n* variables and *m* equality constraints the problem is defined as shown earlier

Maximize (or minimize) $f(\mathbf{X})$, subject to $g_j(\mathbf{X}) = 0, \; j = 1, 2, \ldots, m$

where $\mathbf{X} = \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Bmatrix}$

In this case the Lagrange function, **L**, will have one Lagrange multiplier λ_j for each constraint as

$$L(x_1, x_2, \ldots, x_n, \lambda_1, \lambda_2, \ldots, \lambda_m) = f(\mathbf{X}) + \lambda_1 g_1(\mathbf{X}) + \lambda_2 g_2(\mathbf{X}) + \ldots + \lambda_m g_m(\mathbf{X}) \tag{14}$$

Necessary conditions for a general problem…contd.

L is now a function of *n* + *m* unknowns, $x_1, x_2, \ldots, x_{n,} \lambda_1, \lambda_2, \ldots, \lambda_m$, and the necessary conditions for the problem defined above are given by

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i}(\mathbf{X}) + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i}(\mathbf{X}) = 0, \qquad i = 1,2,\ldots,n \quad j = 1,2,\ldots,m$$

$$\frac{\partial L}{\partial \lambda_j} = g_j(\mathbf{X}) = 0, \qquad j = 1,2,\ldots,m \tag{15}$$

which represent *n* + *m* equations in terms of the *n* + *m* unknowns, x_i and λ_j. The solution to this set of equations gives us

$$\mathbf{X} = \begin{Bmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_n^* \end{Bmatrix} \quad \text{and} \quad \lambda^* = \begin{Bmatrix} \lambda_1^* \\ \lambda_2^* \\ \vdots \\ \lambda_m^* \end{Bmatrix} \tag{16}$$

Sufficient conditions for a general problem

A sufficient condition for $f(\mathbf{X})$ to have a relative minimum at $\mathbf{X}^*$ is that each root of the polynomial in ϵ, defined by the following determinant equation be positive.

$$\begin{vmatrix} L_{11}-\epsilon & L_{12} & \cdots & L_{1n} & g_{11} & g_{21} & \cdots & g_{m1} \\ L_{21} & L_{22}-\epsilon & & L_{2n} & g_{12} & g_{22} & & g_{m2} \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ L_{n1} & L_{n2} & \cdots & L_{nn}-\epsilon & g_{1n} & g_{2n} & \cdots & g_{mn} \\ & & & & & & & \\ g_{11} & g_{12} & \cdots & g_{1n} & 0 & \cdots & \cdots & 0 \\ g_{21} & g_{22} & & g_{2n} & \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots & \vdots & & & \vdots \\ g_{m1} & g_{m2} & \cdots & g_{mn} & 0 & \cdots & \cdots & 0 \end{vmatrix} = 0 \qquad (17)$$

Sufficient conditions for a general problem…contd.

where

$$L_{ij} = \frac{\partial^2 L}{\partial x_i \partial x_j}(\mathbf{X}^*, \lambda^*), \qquad \text{for } i = 1, 2, \ldots, n \quad and \quad j = 1, 2, \ldots, m$$

$$g_{pq} = \frac{\partial g_p}{\partial x_q}(\mathbf{X}^*), \qquad \text{where } p = 1, 2, \ldots, m \text{ and } q = 1, 2, \ldots, n \tag{18}$$

Similarly, a sufficient condition for $f(\mathbf{X})$ to have a relative maximum at $\mathbf{X}^*$ is that each root of the polynomial in Є, defined by equation (17) be negative.

If equation (17), on solving yields roots, some of which are positive and others negative, then the point $\mathbf{X}^*$ is neither a maximum nor a minimum.

Example

Minimize $f(\mathbf{X}) = -3x_1^2 - 6x_1x_2 - 5x_2^2 + 7x_1 + 5x_2$, Subject to $x_1 + x_2 = 5$

Solution

$$g_1(\mathbf{X}) = x_1 + x_2 - 5 = 0$$

$$\mathbf{L}(x_1, x_2, \ldots, x_n, \lambda_1, \lambda_2, \ldots, \lambda_m) = f(\mathbf{X}) + \lambda_1 g_1(\mathbf{X}) + \lambda_2 g_2(\mathbf{X}) + \ldots + \lambda_m g_m(\mathbf{X})$$

with $n = 2$ and $m = 1$

$$\mathbf{L} = -3x_1^2 - 6x_1x_2 - 5x_2^2 + 7x_1 + 5x_2 + \lambda_1(x_1 + x_2 - 5)$$

$$\frac{\partial \mathbf{L}}{\partial x_1} = -6x_1 - 6x_2 + 7 + \lambda_1 = 0$$

$$=> x_1 + x_2 = \frac{1}{6}(7 + \lambda_1)$$

or $\lambda_1 = 23$

$$\frac{\partial \mathbf{L}}{\partial x_2} = -6x_1 - 10x_2 + 5 + \lambda_1 = 0$$

$$=> 3x_1 + 5x_2 = \frac{1}{2}(5 + \lambda_1)$$

$$=> 3(x_1 + x_2) + 2x_2 = \frac{1}{2}(5 + \lambda_1)$$

$$x_2 = \frac{-1}{2}$$

$$x_1 = \frac{11}{2}$$

$$\mathbf{X}^* = \left[\frac{-1}{2}, \frac{11}{2}\right]; \boldsymbol{\lambda}^* = [23]$$

$$\begin{pmatrix} L_{11} - \in & L_{12} & g_{11} \\ L_{21} & L_{22} - \in & g_{21} \\ g_{11} & g_{12} & 0 \end{pmatrix} = 0$$

$$L_{11} = \left.\frac{\partial^2 \mathbf{L}}{\partial x_1^2}\right|_{(\mathbf{X}^*,\boldsymbol{\lambda}^*)} = -6$$

$$L_{12} = L_{21} = \left.\frac{\partial^2 \mathbf{L}}{\partial x_1 \partial x_2}\right|_{(\mathbf{X}^*,\boldsymbol{\lambda}^*)} = -6$$

$$L_{22} = \left.\frac{\partial^2 \mathbf{L}}{\partial x_2^2}\right|_{(\mathbf{X}^*,\boldsymbol{\lambda}^*)} = -10$$

$$g_{11} = \left.\frac{\partial g_1}{\partial x_1}\right|_{(\mathbf{X}^*,\boldsymbol{\lambda}^*)} = 1$$

$$g_{12} = g_{21} = \left.\frac{\partial g_1}{\partial x_2}\right|_{(\mathbf{X}^*,\boldsymbol{\lambda}^*)} = 1$$

The determinant becomes

$$\begin{pmatrix} -6-\in & -6 & 1 \\ -6 & -10-\in & 1 \\ 1 & 1 & 0 \end{pmatrix} = 0$$

$$(-6-\in)[-1]-(-6)[-1]+1[-6+10+\in]=0$$

$$=> \in = -2$$

Since $\in$ is negative $\mathbf{X}^*, \boldsymbol{\lambda}^*$ correspond to a maximum.

Thank you